

THIS WEEK

EDITORIALS

FAIR TRIALS Japan needs an independent scrutineer to check on universities **p.246**

WORLD VIEW Why the Science Media Centre will not fly in the United States **p.247**



DOUBLE-EDGED On the dual purpose of the saw of the sawfish **p.248**

A Russian renaissance?

Vladimir Putin's promise to increase research spending is welcome – but his country's scientific system needs a complete overhaul.

The plight of Russian science was far from the minds of most of the protesters filling Moscow's streets after the disputed parliamentary poll in December and the equally contested presidential elections on 4 March. The protests, the largest in the country for 20 years, reflect the alienation of Russia's intelligentsia and rising urban middle-class from its inbred political leadership, and the autocratic tendencies scarcely veiled behind its assertions of democracy. Yet the crowds' grievances do mirror some of the problems that are keeping Russian science down.

There is no doubt that the scientific system seriously lacks money. Vladimir Putin's pre-election promise to pump funding into basic science and innovation (see page 253) as part of his plan to modernize and diversify Russia's stagnating economy, which is heavily dependent on energy exports, is welcome. His spending promises — and his opponents' general lack of interest in science — may explain why many Russian academics and working scientists voted to return Putin to the presidency.

But Russian science is undergoing a crisis that runs deeper than money: the country's publication output has declined in the past decade despite a slight recovery of the still-meagre public science budgets. More than 20 years after the end of the Soviet Union, Russia's large research community is fundamentally divided over how it would like the domestic science funding system to operate.

Russian scientists of all ages — not just those who have come of age since the fall of the Soviet Union — are sincerely inclined towards liberal democracy and open intellectual discourse. They believe that scientific systems based on peer-reviewed grant applications, although fiercely competitive, are preferable to any form of prescribed knowledge production. But an influential class of old-school academicians and science administrators would rather return to communist secret practices than accept the norms and standards of what they see as Western science. Peer review? Project evaluation? Perish the thought.

It is mostly the latter group — together with Russia's omnipresent Federal Security Service, formerly run by Putin — that has created the stifling bureaucracy and perplexing jungle of regulations and restrictions that many aspiring Russian scientists, let alone their foreign collaborators, have learned to hate. The travel and security restrictions that some Russian physicists are currently protesting against are just the most recent example of the harassment to which the apparatchiks subject scientists (see *Nature* <http://dx.doi.org/10.1038/nature.2012.9921>; 2012). Worse, favouritism and corruption pervade science and higher education — there is bribery from the lowest levels, with students paying teachers in exchange for passing exams, to the highest.

If science is to have a constructive role in shaping Russia's future, Putin must tackle these problems as forcefully as possible. Economists say that a key test of his leadership will be how far he is prepared to go to reform the economy; his agenda should also include kick-starting overdue scientific reform. Rather than relying on the advice of an

exclusive inner circle of buddies and dignitaries, as he has in the past, Putin should set up a truly independent scientific advisory council, ideally involving foreign scientists, to guide him through the necessary changes. Russia's partnership with the Massachusetts Institute of Technology in Cambridge, which will help to set up a new research

"Putin should set up a truly independent science advisory council."

university in Skolkovo, near Moscow, is a first step in that direction. The creation of a well-funded granting agency for university research, with transparent and fair procedures, would be a perfect sequel.

Russia has its own rich traditions and intellectual culture. It need not imitate the scientific landscape of any foreign nation, but the blindness of parts of its scientific establishment to the necessity of change borders on negligence. It is time for that establishment to throw overboard the stubborn obscurantism that stands in the way of a renaissance of Russia's proud scientific past. ■

The shared burden

A proposed change to Germany's constitution is needed for the future health of the universities.

It is hard to find anyone who can offer a sensible explanation for why the 16 German states signed up to a reform in 2006 that prevented the federal government from directly injecting cash into their universities. Federal funds had previously been used to share the considerable burden of financing infrastructure and large equipment in universities, for which the states have general responsibility. The reform made such sharing impossible. And that left universities, particularly those in poorer states, to deteriorate in quality.

Both federal and state governments now agree that the situation must be reversed — urgently, given the relative underfunding of German universities. A proposal on 4 March by education and research minister Annette Schavan aims to do just that.

But the required (two-word) change to the German constitution is opposed by some political parties and certain states. Opponents say that the reform should not be limited to universities, and should allow the federal government to co-finance schools too — a demand that Schavan cannot accept.

To understand the situation, one has to look back to Germany's 1949 post-war constitution, which was designed to ensure that a dictator could never again seize centralized power. This involved the creation of a highly federalized country with politically strong state governments.

In particular, responsibility for education, the most sensitive political issue, was placed exclusively in the hands of the states. The constitution allowed the federal government to offer only short-term project money to universities.

This admirable philosophy was not without problems. Universities in the more cash-strapped states, such as Saarland in the west and Mecklenburg-Vorpommern in the east, could not compete with universities in rich states such as Bavaria and Baden-Württemberg in the south. And frustrated successive federal governments were unable to exert political influence to force universities to raise their game and become competitive. Despite Germany's general wealth, barely half a dozen of its universities regularly appear in the Shanghai Jiao Tong University ranking of the world's top 100 universities. The two most successful ones — the Technical University of Munich (now placed forty-seventh) and the Ludwig Maximilians University of Munich (placed fifty-fourth) — are in Bavaria. The University of Heidelberg in Baden-Württemberg comes in third, at sixty-second.

For decades, the federal and state governments have colluded to make exceptions to the strict federalization of universities in ways — such as in infrastructure — that did not allow the federal government to exert political influence. A 2005 agreement to create the multibillion-euro Excellence Initiative — which will expire in 2017 after two rounds — was something of an exception. The competition for graduate schools, research clusters and the coveted title of 'elite university' indirectly allowed the federal government to exert its political will.

So why the 2006 decision to stop money flowing to universities from the federal government? Probably because the 2006 federal reform was a much wider-ranging attempt to clarify responsibilities

between governments on both levels that had, over the decades, been rendered fuzzy by the introduction of innumerable exceptions to the divisions laid out in the 1949 constitution. Universities were a tiny part of this. Although experts saw the problems immediately, they could not persuade politicians to delay a highly complicated agreement to negotiate an apparent detail.

The federal government could do even more to steer academic science in Germany towards sustained competitiveness.

its inclusion in the German Council of Science and Humanities, a high-level committee of scientists and politicians.

The debate over schools could involve years of painful negotiation, and German universities should not be made to wait. If the Social Democrats and Greens drop their opposition to the two-word addition to the constitution now, that change could be in place by March 2013. They should do so. The federal government could then, for example, directly ensure the survival beyond 2017 of successful institutes created in the Excellence Initiative, even if the states in which they are located cannot give them permanent financing. And the government could do even more to steer academic science in Germany towards sustained competitiveness — not to usurp the states' power but to give them financial breathing space. ■

Final say

Ongoing controversy over work at Japan's Tohoku University must be resolved.

It is not easy for universities to investigate their own researchers. But as a case at Tohoku University shows, it is especially difficult in Japan — particularly when the scientist concerned is the university's president.

When concerns were raised in 2007 about irregularities in the research of Akihisa Inoue, who specializes in making metallic glasses, the university formed a committee to look into the matter. The committee determined that there was no case to answer and that the matter need not be investigated formally. But the decision satisfied few of his doubters, who criticized the committee for being heavily populated by university administrators whom Inoue had appointed to their posts (but not to the committee). The university says that there was no conflict of interest.

Four years later, and with claims of problems in Inoue's papers accumulating, the university formed another committee. This time it got an eminent outsider, Akito Arima, former president of Tokyo University, parliamentarian and minister of education, to lead. The committee focused on retractions of six papers on which Inoue had been an author.

The committee concluded that the duplicate submissions that led to the retractions were an example of poor, but somewhat accepted, practice in the field of materials science. Duplicate publications, the report said, were bad, but "different from practices such as FFP [falsification, fabrication and plagiarism], which clearly have a negative impact on research." In a formal statement, Inoue apologized and said that the university would quickly implement the report's recommendations.

The judgement was surprisingly light. Six retractions is a serious

matter, although they may simply indicate that Inoue — who has more than 2,500 papers to his name — was trying to move too fast. As Inoue has told *Nature*, his laboratory did not always check its papers to ensure that its research was cited properly.

Then, in April 2011, the Japan Science and Technology Agency (JST) formed a committee to find out whether the retractions, corrections and other allegations had undermined the results of a project it had funded between 1997 and 2002. It found no reason to question the main findings, but concluded that 5 of 27 high-impact papers were duplicated in other papers by Inoue's laboratory. But neither of these latest committees, as both acknowledge, investigated the more serious allegations against Inoue concerning irreproducibility of results (see *Nature* 470, 446–447; 2011). Inoue denies any manipulation of data, and there is no evidence that he has committed any scientific misconduct.

The JST, in accordance with its policies, has now requested that the university look into the allegations. The university says that it will do so, but will give no details about the composition of the committee, or about the investigation's mission, scope or expected date of completion.

Unlike countries such as the United States, Japan has no final arbiter for such issues. The Office of Research Integrity in Rockville, Maryland, for instance, as well as many funding agencies, can aggressively investigate allegations. Even if they don't satisfy everyone, their decisions are final, and for the most part fair.

The primary responsibility for investigating should always remain with the university, research institute or relevant scientific society. But Japan needs a procedure to oversee the committees that make these judgements. The authors of the JST report suggest that the Science Council of Japan (if given the necessary funding to repeat experiments) could fit this role. Such a body seems worthwhile, and

it could very well have closed the arguments at Tohoku University that refuse to die even after nearly five years. Japan's policy-makers should use the situation at Tohoku to create a mechanism to handle situations such as these. ■

► NATURE.COM

To comment online,
click on Editorials at:
go.nature.com/xhunq



Two nations divided by a common purpose

Plans to replicate Britain's Science Media Centre in the United States are fraught with danger, warns Colin Macilwain.

Marks & Spencer. Oasis. Football (proper football) for the first hundred years. The list of great British ideas to fall flat on their faces in the United States is a lengthy one. It is a roll-call of infamy that sprang to mind last month, when plans emerged to establish a Science Media Centre in the United States.

The original Science Media Centre (SMC) began in London in 2002. It offers the media a clearing house for scientific briefings and packaged quotes from scientists. On its own terms, it has been an outstanding success. A combination of factors lies behind this impact: the energy of the centre's staff; the backing of sponsors, including scientific societies, major corporations and most governmental and non-governmental research funders; and a close-knit London media circle.

Science media centres similar, but not identical, to that in London already exist in locales as far apart as Australia, Canada and Japan. But it seems to me, as someone who has worked as a reporter and an editor on both sides of the Atlantic, that there are formidable obstacles to a successful introduction of the concept to the United States.

For a start, the 'problem' regarding science and the public is different in each place. The London SMC was set up because UK scientific leaders were upset that environmentalists had successfully fought the introduction of genetically modified food; they felt that the UK media were too susceptible to environmental scare stories about new technologies.

The main public-relations challenge facing science in the United States is different — bitter social and political division over stem-cell research, global warming, creationism and much else besides. A US Science Media Centre would either avoid these highly partisan issues — and face irrelevance — or step right into them, and take a level of heat that the UK SMC has never experienced.

Second, US journalists, justifiably or not, have higher self-regard than their British counterparts and are likely to take strong issue with the 'churnalism' aspects of the SMC. Under pressure as US reporters may be, they don't want to share 'pooled' quotes.

Third, the entire media picture in the United States is much larger and more diverse. There isn't the direct competition between titles that obliges the main London newspapers to run so many stories every day, in case readers miss something. Broadcasting dwarfs print media in its US reach, and is itself now being eclipsed by a plethora of new media — ranging in scope from an unread Tweet to *The Huffington Post* — as the principal source of news.

Supporters say that the US SMC would reflect this scale and diversity, doing most of its work

online. But it is hard to see how a digital SMC would differentiate itself from other interest groups.

Finally, the top echelons of the US media — what Newt Gingrich would call the 'liberal media elite' — are particularly sensitive to the nature of their sources of information. If an organization in Washington or New York, co-funded by government and big business, tried to sell them a line, they would recoil.

The London SMC's narrow approach to risk assessment — if you want to hear about the risks of nuclear power, say, just ask your local nuclear engineer (see *Nature* 471, 549; 2011) — sits happily with the prevalent ethos of British journalism. This was, of course, immortalized by the otherwise-obscure poet Humbert Wolfe: "You cannot hope | to bribe or twist, | thank God! the | British journalist. | But, seeing what | the man will do | unbribed, there's | no occasion to."

Despite the fears of the SMC founders, the British press — led by the BBC, which treats the Confederation of British Industry with the deference the Vatican gets in Rome — is overwhelmingly conservative and pro-business in its outlook. It is quite unperturbed by the fact that SMC sponsors include AstraZeneca, BP, Coca-Cola, L'Oréal, Monsanto, Syngenta (as well as Nature Publishing Group) but not a single environmental non-governmental organization (NGO) or trade union.

Fiona Fox, the SMC's director, says that the centre operates independently of its sponsors and points out that none (except its host, the Wellcome Trust) accounts individually for more than 5% of its income. She adds that no NGOs are involved because it was their public-relations

skills that the founders of the SMC sought to match. But the perception that the environmental group Friends of the Earth constitutes a bigger threat to scientific truth-telling than some of the corporate names on the SMC's sponsorship list is not one the US media would accept.

Some of those considering a US centre share these concerns. They think that their funding model will have to rely on charitable trusts, not companies or government agencies. And they see the US SMC as a source of background information and advice to help reporters to get out and do their jobs: talking to their own sources and obtaining their own quotes. That agenda would be less ambitious — and less fraught with danger.

This spring, reporters and funders in the United States will be consulted on whether the Science Media Centre can mimic the transatlantic success of a certain four-piece band from Liverpool. I fear, instead, that the fate of Robbie Williams awaits. Who, you ask? Well, exactly. ■

US SCIENCE FACES A BITTER SOCIAL AND POLITICAL DIVISION OVER STEM CELLS, GLOBAL WARMING AND CREATIONISM.

© NATURE.COM
Discuss this article
online at:
go.nature.com/klnuna

Colin Macilwain writes about science policy from Edinburgh.
e-mail: cfnworldview@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

GEOPHYSICS

Rain slows seismic waves

The recent occurrence of earthquakes and even heavy rainfall reduce the near-surface velocity of a type of seismic wave responsible for the worst damage during earthquakes.

Nori Nakata and Roel Snieder at the Colorado School of Mines in Golden studied these 'shear' waves using seismic interferometry, which measures how the waves propagate from depth to the surface. The authors analysed the average velocities of shear waves in nearly 112,000 quakes — most of them minor — recorded between 2000 and 2010 by Japan's network of seismic stations. They found that, three months after major earthquakes, wave velocities recorded at nearby stations were 3–4% lower than normal.

In southern Japan, shear-wave propagation was also slower during rainy summers than drier seasons, possibly because of the increase in fluid pressure that occurs when the ground is infiltrated by rainwater.

J. Geophys. Res. <http://dx.doi.org/10.1029/2011JB008595> (2012)

PHYSIOLOGY

Sawfish snouts sense and stab

The serrated snouts of sawfish have an unusual double role: the fish use them to both detect and disable prey.

A number of fish species have elongated 'noses' that they use either to detect or to attack food. In studies of captive juvenile freshwater sawfish (*Pristis microdon*), Barbara



M. CRISP/AUSTRALIAN NATL UNIV.

PLANT EVOLUTION

Outback palms were planted

The Australian outback's only palm trees, thought to be living fossils of the ancient Gondwanan rainforest that covered the region 15 million years ago, may in fact be descended from transplants carried by humans.

Toshiaki Kondo at Hiroshima University and his colleagues compared DNA samples from four populations of the outback's *Livistona mariae*, which give their name to the Palm Valley Oasis, with those of other Australian palms. They found that the species is genetically almost identical to

its northern relation, *Livistona rigida* — meaning that the two are not distinct species. Genetic analysis revealed that the populations started to diverge between 7,000 and 31,000 years ago, a timeline that overlaps with intermittent human migrations from northern to central Australia.

The authors suggest that Aboriginal people may have brought the palm seeds with them as a food resource or raw material.

Proc. R. Soc. B <http://dx.doi.org/10.1098/rspb.2012.0103> (2012)

ORGANIC ELECTRONICS

Sterilization-ready transistors

The striking flexibility of organic electronics means that they hold immense promise for use in biomedical devices such as catheters. But most organic materials melt or degrade when heated to the temperatures needed to sterilize materials, as required for medical use.

Takeo Someya at the University of Tokyo and his colleagues have built a flexible organic thin-film transistor that is heat resistant and works at the low voltages that might

be used in the human body. They heated the device, which consists of a thiophene-based aromatic active layer with aluminium oxide and alkyl phosphonic acid insulating layers, up to 150 °C and found that its electrical characteristics remained stable.

Nature Commun. 3, 723 (2012)

CANCER GENETICS

Many genomes in one tumour

Separate parts of a single tumour can contain distinct genetic mutations, and clinicians might suggest



contradicting prognoses or treatments on the basis of biopsies from different areas. Charles Swanton at Cancer Research UK's London Research Institute and his colleagues analysed this diversity by reconstructing the evolution of four kidney cancers, sampling several parts of each primary tumour, as well as secondary ones from organs to which the cancer had spread.

In one patient, only one-third of the total mutations were found in all the samples, and one-quarter were unique to just one. This has implications for biomarker analysis and cancer-genome studies, which rely on single biopsies. It might also explain why many cancer treatments eventually stop working: they affect only part of a tumour, allowing the remainder to bounce back.

N. Engl. J. Med. 366, 883–892 (2012)

PLANETARY SCIENCE

Venusian hot flow anomalies

A giant aneurysm of plasma distorted Venus's atmosphere on 22 March 2008, a space-weather event known as a hot flow anomaly (HFA). Such events occur when electric fields associated with the Sun's solar wind create a fast-growing bulge that is filled with hot ions from the bow shock, a boundary between the solar wind and the planet.

Venus joins Earth, Saturn and Mars as planets where HFAs have been observed, say Glyn Collinson at the Goddard Space Flight Center in Greenbelt, Maryland, and his colleagues, who used data from the European Space Agency's Venus Express satellite to identify the 2008 event.

On Earth, HFAs can cause auroras. The authors suggest that HFAs would be even more disruptive to Venus's atmosphere because the planet lacks a magnetic field, meaning that the bow shock, and so the HFA, would be much closer in.

J. Geophys. Res. <http://dx.doi.org/10.1029/2011JA017277> (2012)

IMMUNOLOGY

Controlling natural killers

Immune cells long considered lowly foot soldiers on the front line against bacteria and fungi turn out to have a commanding role. Researchers report that neutrophils, the most common white blood cells in humans, control the activity of another antimicrobial white blood cell — the natural killer cell.

Sophie Ugolini and Eric Vivier of Aix-Marseilles University in France and their colleagues isolated a mutant mouse whose natural killer cells responded poorly to stimulation. The mutant, they found, lacked mature neutrophils. Without neutrophils to guide them, mouse natural killer cells failed to mature or function properly. The same was true of natural killer cells in human patients with low neutrophil counts, a condition known as neutropenia.

J. Exp. Med. <http://dx.doi.org/10.1084/jem.20111908> (2012)

PALAEOECOLOGY

What killed the big beasts?

During the past 100,000 years, many of Earth's largest animals — including ground sloths, mammoths, woolly rhinos and sabre-toothed cats — became extinct. Scientists have debated why for decades, with climate change and hunting by humans the chief suspects. Graham Prescott, David Williams and their group at the University of Cambridge, UK, have created a model of unprecedented geographical breadth — and concluded that it took both factors to seal the beasts' fate.

The researchers modelled extinctions in North America, South America, Palaeoarctic Eurasia, Australia and New Zealand, running simulations with climate data from ice cores and thousands of plausible combinations of human-arrival and species-extinction

COMMUNITY CHOICE

The most viewed papers in science

GENETICS

James Crow on sabotaging sperm

★ HIGHLY READ
on plosgenetics.org 5 Feb–6 Mar

A recent paper identifies a cause for sperm with puzzlingly high rates of mutation that cause a particular thyroid cancer syndrome.

But the Perspective that accompanied it was viewed more often than the work itself, perhaps because its author, James Crow, a distinguished geneticist at the University of Wisconsin, Madison, died in January, aged 95.

Those who read Crow's last paper found the cause of the high mutation rate summarized in his clear and pleasant prose. Rather than dividing asymmetrically — to produce one sperm and one proto-sperm cell like the parent — some cells 'cheat' and produce two proto-sperm cells instead. Each of these can divide asymmetrically, doubling the cheat's number of descendants and creating clusters of mutations linked to the effect.

PLoS Genet. 8, e1002535 (2012)

For an obituary of James Crow, see go.nature.com/ty6xi5

times, reflecting the large uncertainty in both these estimates. The recipe that best predicted the pattern of extinctions included both climatic and human ingredients.

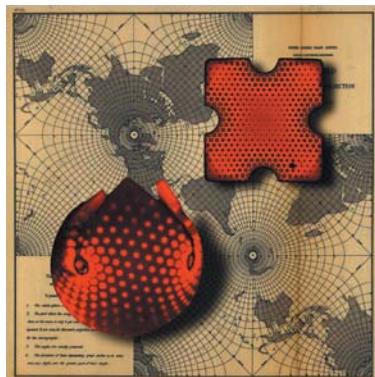
Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1113875109> (2012)

MATERIALS

A new way to 3D shapes

Materials that can change their shape on demand have many applications, and, thanks to a team in the United States, an extra means of production. Christian Santangelo, Ryan Hayward and their colleagues at the University of Massachusetts at Amherst report that they have devised a way to produce polymer sheets that can transform into predetermined three-dimensional shapes and back again.

The researchers created gel sheets from polymers of *N*-isopropylacrylamide and benzophenone, a molecule that creates crosslinking between polymers when exposed to ultraviolet (UV) light. These sheets swell in



water, but shrink when heated.

The team patterned the gel using a mask and an initial low dose of UV to define the polymer's overall shape, then used a higher dose through a different mask to create highly crosslinked regions in which swelling is reduced. By allowing more swelling in some areas than others, the authors created objects that can switch between sheets and spheres when heated and cooled (pictured), as well as others that can reversibly form caps, cones and more-complex surfaces.

Science 335, 1201–1205 (2012)

NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Telescope rivals

A scientific panel has narrowly recommended South Africa over Australia as the best site for the proposed Square Kilometre Array (SKA), a US\$2.1-billion international radio telescope (see *Nature* **480**, 308–309; 2011). But the project's voting member states have yet to make the final call on where the telescope will go. Despite the SKA Site Advisory Committee's recommendation, both sites are still in contention, a source told *Nature*. The final choice could come as soon as 4 April. See go.nature.com/gwfejy for more.

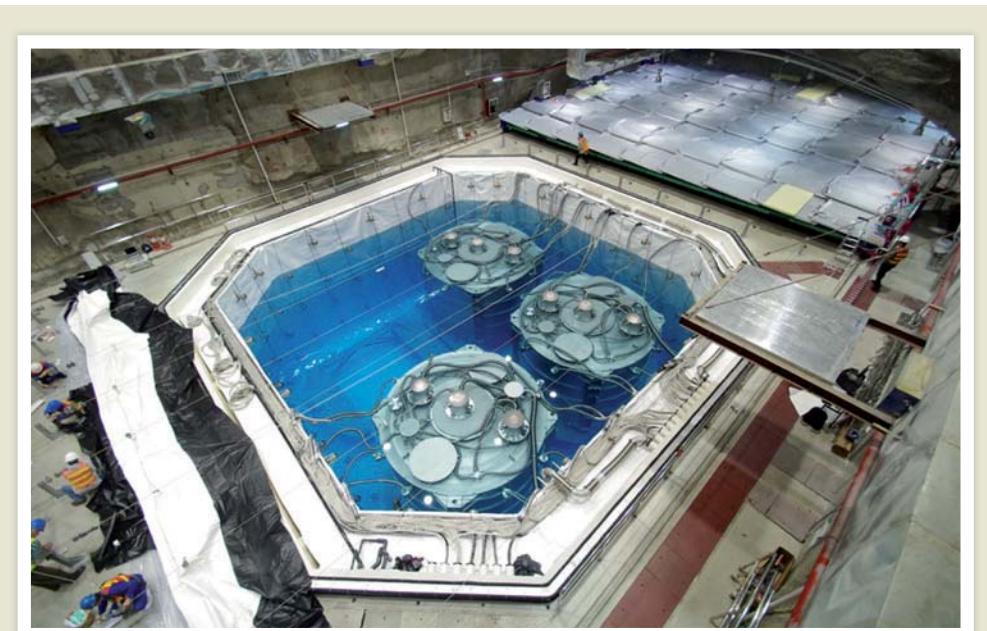
Tevatron ♥ Higgs

Before it was shut down last year, the Tevatron particle collider at Fermilab in Batavia, Illinois, gathered data from proton–antiproton collisions that seem to support the existence of the elusive Higgs boson predicted by the standard model of physics. The results, announced on 7 March at the Moriond conference in La Thuile, Italy, do not reach a level that physicists would consider statistically significant. But they are consistent with reports in 2011 of a Higgs particle with a mass of around 125 gigaelectronvolts, from experiments at the Large Hadron Collider at CERN, Europe's high-energy physics lab near Geneva, Switzerland. See go.nature.com/cd3rma for more.

POLICY

FDA improvement

Scientists who work for the US Food and Drug Administration have greater respect for their leaders and more confidence in the agency's future than they did five years ago, according to a survey of 997 researchers



XINHUA/PHOTOSHOT

Success for China's neutrino test

The final parameter needed to quantify the rate at which neutrinos change ('oscillate') from one form to another has been precisely measured by a neutrino-detection experiment in southern China. The Daya Bay Reactor Neutrino Experiment (pictured), on the Dapeng Peninsula near Hong Kong, monitors electron antineutrinos produced by nearby

nuclear power plants. Its measurement, reported on 8 March, pips efforts by reactors in Japan, France and the United States. It also validates experiments that plan to compare the rates of neutrino and antineutrino oscillations, to test the symmetry between matter and antimatter. See go.nature.com/9yssr1 for more.

(see go.nature.com/cv5gi6) published on 7 March by the Union of Concerned Scientists in Washington DC. But researchers are still uneasy about their rights to publish work that might be considered contentious, and to talk to the press and public about their findings, the survey showed. They were also worried about corporate and political influence on the agency's decisions.

Antarctic repair

The Brazilian government last week laid out plans to rebuild and modernize the country's Antarctic research station, which was mostly destroyed by a fire on 25 February.

No money has yet been allocated for construction, but the science ministry has committed more than US\$5.5 million in new grants to rebuild research projects, and funds to replace lost scientific equipment. See go.nature.com/wluzf1 for more.

Oil-spill restoration

Restoration efforts in the Gulf of Mexico are likely to receive around 80% of the pollution fines that the US government collects from oil giant BP as a result of the 2010 oil spill in the region. On 8 March, the US Senate passed a bill (the RESTORE the Gulf Coast Act) that would approve this

allocation of funds, which would otherwise go to the US Treasury. Federal fines, which are separate from private settlements, are expected to range from US\$5 billion to \$20 billion. The bill is now attached to a transportation bill that awaits passage through Congress.

Water scarcity

The world needs to rethink its policies on managing freshwater resources, which are being severely threatened by climate change and a growing population, warned the United Nations Educational, Scientific and Cultural Organization on 12 March. The Paris-based

UNIV. CALIFORNIA, IRVINE/AP
organization launched its triennial *World Water Development Report* at the start of a meeting to discuss the problem in Marseilles, France. See page 256 for more.

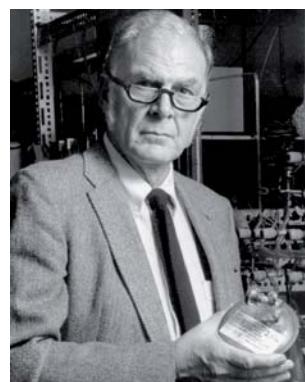
PEOPLE

Brain prize

Two women who studied genes involved in hearing share this year's Brain Prize, a €1-million (US\$1.3-million) annual award given to European neuroscientists since 2011 by the Grete Lundbeck European Brain Research Foundation in Copenhagen. Christine Petit of the College of France and the Pasteur Institute in Paris identified genes underlying different types of deafness. Her co-winner, Karen Steel of the Wellcome Trust Sanger Institute in Hinxton, UK, mapped key mouse genes involved in the function of cochlear hair cells, which pass sounds to the brain.

Nobel chemist dies

Frank Sherwood ('Sherry') Rowland (pictured), co-winner of the 1995 Nobel Prize in Chemistry, died on 10 March aged 84. One of the world's most famous climate scientists, in 1974 he reported with Mario Molina that chlorofluorocarbons (CFCs) were helping to destroy the ozone layer. Rowland advocated



banning the chemicals; his work led to the Montreal Protocol in 1987. He continued to study atmospheric pollution until his death, and won the Nobel prize for his ozone research. Rowland was also a founder member of the University of California, Irvine. See go.nature.com/hpmb8j for more.

Ocean dives

Mixing showbiz, adventure and science, movie director James Cameron says that he intends this month to dive solo to the bottom of the 11-kilometre-deep Mariana Trench in the Pacific Ocean. On 8 March, Cameron, together with the US National Geographic Society, announced his plans — as part of a documentary — to take the first manned visit to the trench since the *Trieste* submersible visited the Challenger Deep region in 1960. British billionaire entrepreneur

Richard Branson is also hoping to reach the trench, although his Virgin Oceanic mission is suffering delays.

Retraction spree

A Japanese anaesthesiologist who published work without ethics approval from an institutional review board has been fired by his university and had eight papers retracted, with many of his others under suspicion. Toho University in Tokyo said on 6 March that it had dismissed Yoshitaka Fujii on 29 February after an investigation. On 7 March, the journal *Anesthesia & Analgesia* published an expression of concern about 24 other papers by Fujii, and said that its editors had spent two years following up allegations of research misconduct — one made as early as 2000 but not followed up until recently.

SEVEN DAYS THIS WEEK

COMING UP

19–23 MARCH

The latest developments in nuclear-powered spaceflight are discussed at the Lunar and Planetary Science Conference in The Woodlands, Texas. It marks the 50th anniversary of the launch of the first nuclear-powered satellite.

go.nature.com/inloal

20–21 MARCH

Chemists, biologists and regulators meet in Brussels to discuss the transition to animal-free toxicity testing.

go.nature.com/ptvih2

BUSINESS

Indian generics

In a landmark ruling, the Indian Patent Office has allowed a domestic firm to sell a generic version of a patent-protected anti-cancer drug so that the treatment is affordable. Under a compulsory licence granted on 12 March, German pharmaceutical giant Bayer must allow Natco Pharma in Hyderabad to make and sell its patented drug Nexavar

(sorafenib) in return for 6% royalties on net sales. Natco will charge patients 8,800 rupees (US\$176) a month; Bayer was charging 280,000 rupees. Bayer may challenge the licence, which is permitted under the World Trade Organization's Agreement on Trade-Related Aspects of Intellectual Property Rights.

EVENTS

Space weather

A strong geomagnetic storm hit Earth on 9 March last week, three days after the Sun spat out a large solar flare. No serious damage was reported, although such solar flares can cause radio blackouts and affect satellite navigation systems. About 200 geomagnetic storms of this intensity (G3) are expected for every 11-year solar cycle, according to the US government's Space Weather Prediction Center in Boulder, Colorado. As the Sun approaches its solar maximum in 2013, more are expected.

NATURE.COM

For daily news updates see:
www.nature.com/news

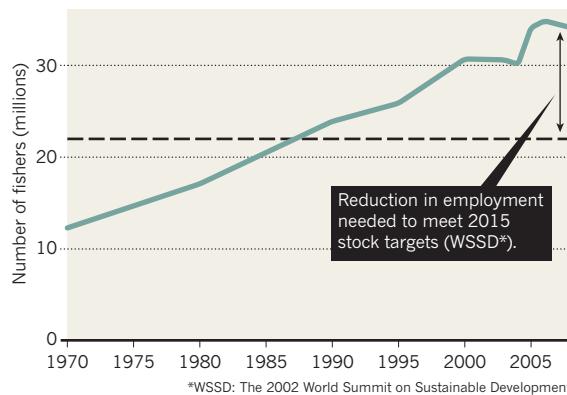
TREND WATCH

SOURCE: FISH & FISHERIES

To reduce pressure on fish stocks to a sustainable level by 2015, the worldwide fishing effort must be scaled back by 36–43% of 2008 levels, according to an analysis by researchers at the Food and Agriculture Organization of the United Nations in Rome (Y. Ye *et al.* *Fish Fish.* <http://doi.org/hq8>; 2012). That would mean cutting the jobs of between 12 million and 15 million fishers (see chart), and would cost up to US\$358 billion — but it would increase production in the long term.

TOO MANY FISHERS, NOT ENOUGH FISH

To meet 2015 sustainable-development targets for healthy fish stocks, between 12 million and 15 million fishers will have to lose their jobs.



NEWS IN FOCUS

CLIMATE CHANGE Modellers look to the past to prepare for UN report **p.254**

SPACE An X-ray mission to reveal the black hearts of galaxies **p.255**

RESOURCES Putting the world's water crisis on the map for Rio **p.256**

CROPS The quest to create a caffeine-free coffee bean **p.264**

S. MORDOVETS/GETTY



Vladimir Putin is proposing to set up world-class institutes and broaden distribution of research funding.

POLITICS

Putin promises science boost

But Russian researchers are sceptical of ambitious schemes.

BY QUIRIN SCHIERMEIER

As a teary-eyed Vladimir Putin celebrated a triumphant win in the Russian presidential elections last week, the country's scientists expressed mixed feelings about the implications for research.

Putin, leader of the United Russia party and

currently prime minister, will in May replace Dmitry Medvedev as president — a post that he previously held in 2000–08. Science had only a marginal role in the election campaign, but Putin outlined his ambitions for research in the newspaper *Vedomosti* on 30 January.

His government, he wrote, will aim to establish several "world-class" research universities

by 2020. Many rectors and chancellors of Russian universities are delighted, and campaigned for Putin. Indeed, there are high hopes for the planned Skolkovo Institute of Science and Technology, a graduate research university set to be established in a science city outside Moscow, in partnership with the Massachusetts Institute of Technology in Cambridge. But critics say that Putin's goal is out of reach, given that there are currently no Russian institutions in the top ranks of world university league tables.

Putin also promised a substantial increase in public funding of basic and applied research. The Russian science funding system, often criticized for cronyism and a lack of fair grant-assessment procedures, must become more transparent and open to genuine competition, he wrote. Some researchers are cautiously optimistic, acknowledging that conditions have improved in recent years, thanks to modest funding and salary increases. "I do think that science plays a serious role in Putin's thinking about Russia's future," says Raul Gainetdinov, a neuroscientist at the Italian Institute of Technology in Genoa, who left Russia in 1996. "If opportunities at home were to really improve, I'd be more than happy to return."

But although Russian research budgets have increased since their near-collapse in the 1990s, the country's output of scientific papers has declined since 2000, putting it below India, Australia and Canada, according to a 2010 analysis by global data-provider Thomson Reuters. Funding opportunities are poor, with only a select few scientists able to benefit from flagship projects such as the space programme, or the 'mega-grants' launched in 2010 to attract leading researchers to the country (see *Nature* **465**, 858; 2010). For the majority of Russian scientists, project grants are almost impossible to get.

"I receive a salary from the Russian Academy of Sciences, but there wasn't any money available last year for buying equipment, doing field research or travelling to international conferences," says Leo Borkin, a herpetologist at the academy's prestigious Zoological Institute in St Petersburg and former head of the St Petersburg Association of Scientists and Scholars. "Young scientists here have basically no opportunity to make ties with colleagues from North America, Europe or Asia," he says. "It's a very sad and stupid situation."

Putin has promised to increase the ►

► amount of grant money distributed by funding agencies from around 15 billion roubles (US\$500 million) a year to 25 billion roubles by 2018. The average size of grants awarded by the Russian Foundation for Basic Research, for example, currently some 350,000 roubles (US\$12,000) per year, will be "made comparable to Western grants", he wrote.

Researchers welcome more cash, of course, but the plan implies that the number of grants might actually go down, notes Konstantin Severinov, who runs independent groups at the Russian Academy of Sciences (RAS) institutes for molecular genetics and gene biology in Moscow, and Rutgers University

in New Brunswick, New Jersey.

Putin also intends to break the long-standing dominance of the academy, which employs around 50,000 scientists at more than 400 research institutes. He plans to redistribute some of the RAS's budget — currently about 50 billion roubles per year — to give other institutions and universities more money, as a key part of a sweeping ten-year science plan that he has asked the academy to develop.

Last month, the RAS administration made a start by asking leading scientists at its institutes to compile lists of research results that they expect to produce by 2030, including cost estimates. "This is just not how science works,"

says Severinov. To him, the unrealistic request is typical of the stifling bureaucracy prevalent in the Russian science system. "There are undeniably some advances in how science is run in this country. Alas, there are lots of missed opportunities as well," he says.

Mikhail Gelfand, a Moscow-based bioinformatician who in December spoke to a 100,000-strong crowd of anti-Putin protesters, believes that, under Putin, a "background of omnipresent bureaucracy and corruption" will continue to hamper any Russian science revival. "It would be naive to expect that science alone somehow could, miraculously, blossom in this atmosphere," he says. ■ **SEE EDITORIAL P.245**

CLIMATE SCIENCE

Forecasters look back in time

As the next IPCC assessment nears, scientists use palaeoclimatic data to hone their models.

BY JEFF TOLLEFSON

It is hard to imagine beaches in northern Greenland or driftwood washing up on islands of Canada's Arctic archipelago, but both were a reality some 6,000 years ago. "At least seasonally, those areas must have been ice-free," says Gavin Schmidt, a climate modeller at NASA's Goddard Institute for Space Studies in New York. The warming was caused primarily by cycles in Earth's orbit, not a spike in greenhouse-gas levels, but climate modellers are revisiting the long-gone era as they prepare to deliver their next global forecast.

The temperate mid-Holocene epoch is one of three episodes included in the World Climate Research Programme's Coupled Model Intercomparison Project, which will form the basis of next year's fifth assessment by the Intergovernmental Panel on Climate Change (IPCC). Simulating palaeoclimates is a challenge, but many scientists say they have enough data from sources such as marine-sediment cores, tree rings and cave deposits to evaluate various aspects of their models. "It's important to check the ability of our climate models to simulate a

different climate, because our future world is also quite different from the present," says Masa Kageyama, a palaeoclimate modeller at the Pierre-Simon Laplace Institute near Paris.

Owing, in part, to past computational limitations, palaeoclimate simulations have never been systematically factored in to the IPCC modelling effort. With a July deadline for submitting papers approaching, members of the IPCC's physical-science working group gathered in Honolulu, Hawaii, for a pair of workshops, held over the past fortnight, to discuss palaeoclimate modelling and the range of other simulations on which their assessment depends. Scientists presented early results from the latest generation of climate models, which take into account factors such as the impact of land use and vegetation, as well as an increasingly sophisticated treatment of atmospheric physics and chemistry. The simulations explore the effects of a broad range of potential future greenhouse-gas concentrations, including an extreme case in which the amount of carbon dioxide rises to more than three times its current level.

Some wondered whether the increased complexity could add to the uncertainties of the new

models and create a larger range of climate projections, but Jerry Meehl, a climate scientist at the National Center for Atmospheric Research in Boulder, Colorado, and one of the workshop organizers, says the spread of projections is "roughly the same" as that produced by the previous generation of models. Ron Stouffer, a cli-

mate researcher at the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey, says that his team's model has already delivered surprises on the increase in atmospheric carbon dioxide levels. "It turns out that land-use changes, right up to about 1950 or even 1970, were as large a player as fossil-fuel emissions were."

in atmospheric carbon dioxide levels. "It turns out that land-use changes, right up to about 1950 or even 1970, were as large a player as fossil-fuel emissions were," he says. "And even today they are not trivial."

More than 20 groups around the world are still processing and uploading modelling data onto a server network that will be accessible to everyone involved in the exercise. So far, more than 2 million files and a petabyte of data have been uploaded. Organizers expect that amount to triple, taking it to roughly 100 times that amassed during the last round of modelling for the IPCC's fourth assessment in 2007. The amount of data is creating new challenges in terms of information management, says Karl Taylor, who works with modellers on experiment design at Lawrence Livermore National Laboratory in Livermore, California. "We're seeing an awful lot of results," Taylor says, "but we're also up against deadlines, and everyone is a bit frantic." ■

MORE ONLINE

TOP STORY



LSD helped problem drinkers battle the bottle

go.nature.com/jytmr

MORE NEWS

- Industry ties remain rife on panels for psychiatry manual go.nature.com/ehoeyt
- Retreating ice leaves invertebrates on the rocks go.nature.com/nooiru
- The 'most important questions' in science policy go.nature.com/awlrsy

ASTROPHYSICS

Spacecraft aims to expose violent hearts of galaxies

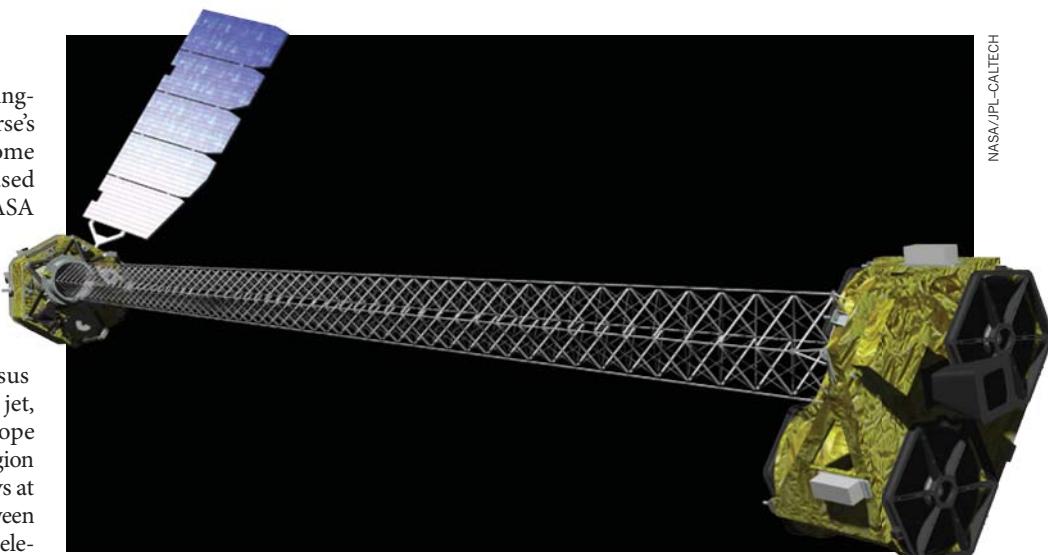
Low-cost mission will tap into the unexplored upper reaches of the X-ray spectrum.

BY ERIC HAND

Who would have thought that a ring-side seat at some of the Universe's most extreme events could come cheap? But by the standards of space-based astronomy, the NuSTAR telescope that NASA plans to launch as early as this month has a modest budget, US\$165 million. Yet it will be sensitive to the high-energy photons produced at the turbulent thresholds of supermassive black holes.

Due to be lofted into orbit by a Pegasus rocket launched in mid-air from a carrier jet, NuSTAR (Nuclear Spectroscopic Telescope Array) is taking aim at an under-explored region of the spectrum. It will detect 'hard' X-rays at 5–80 kiloelectronvolts, an energy range between the softer, lower-energy X-rays probed by telescopes such as the Chandra X-ray Observatory and the γ -rays measured by satellites such as the Fermi telescope. The spacecraft, which features an array of technical innovations, "will be the first mission that can resolve these high-energy X-rays", says Fiona Harrison, an astronomer at the California Institute of Technology in Pasadena and the mission's principal investigator.

Hard X-rays are notoriously difficult to focus, because they tend to penetrate rather than reflect off mirrors, even those coated with dense metals such as gold and iridium. So Harrison and her colleagues have developed a novel material: a sandwich of hundreds of thin



A deployable mast will allow the NuSTAR space telescope to image high-energy X-rays.

metallic films. The thickness of each layer is tuned to reflect photons of a specific energy. Each layer reflects a small number of photons, but they add up to produce a strong reflection. "That's the key technical breakthrough," says Harvey Tananbaum, an astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and director of the Chandra telescope. "Nobody has been there with this kind of sensitivity."

Still, the X-rays are so powerful that they can be reflected only at glancing angles, and the mirrors must therefore be arranged in conical shells, like nested Russian dolls, which funnel the high-energy photons into a detector at the end of the structure. With 133 of these mirror shells in each of two units, NuSTAR will be more sensitive than the Chandra telescope (see 'Going to extremes'), which has only four mirror shells — although, because they are stiffer and more finely polished, Chandra can perceive objects in greater detail.

Grazing-incidence optics means that a long focal length is needed — and, conventionally, a bulky, costly spacecraft. The \$2-billion Chandra could barely squeeze into the cargo bay of the Space Shuttle. NuSTAR, at 10 metres, will be just as long — but it has to fit on the much smaller, cheaper Pegasus launcher. The solution: the telescope is built around a folded-up truss that will grow to full length once NuSTAR is deployed in low-Earth orbit. For astronomers,

the 26 minutes while NuSTAR unpacks itself will be the most anxious phase of the mission.

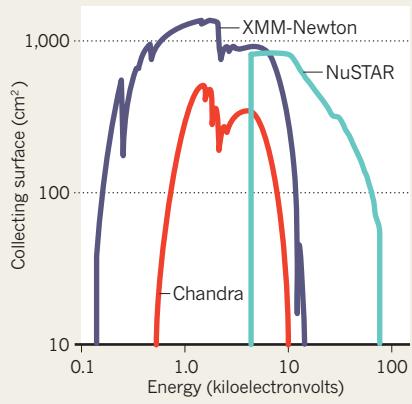
The telescope will focus on active galactic nuclei (AGN), the luminous hearts of distant galaxies, in which highly energized particles fling off X-rays as they whirl around supermassive black holes. Because NuSTAR is sensitive to the most penetrating X-rays, it should be able to spot several hundred new AGN that are shrouded from other detectors by dust and hot gas. That should help astronomers to get a handle on how big and common AGN are, says Harrison, which in turn will help to illuminate a major cosmic puzzle: what role do they have in the growth and evolution of their host galaxies?

The European Space Agency's XMM-Newton mission, another large X-ray telescope, will join forces with NuSTAR to probe AGN. The two spacecraft will look for distortions in the X-ray spectra resulting from the spin of the central black hole. These spin rates, in turn, could offer insight into how the black holes in some AGN grew to masses of billions of Suns — whether by swallowing other AGN in galactic mergers or by steadily accumulating material from the host galaxy. XMM-Newton is devoting 10% of its guaranteed observation programme time to joint observations with NuSTAR.

"It shows how important this mission is to the community," says Norbert Schartel, the project scientist for XMM-Newton at the European Space Astronomy Centre in Madrid. ■

GOING TO EXTREMES

Compared with other X-ray telescopes, NuSTAR has a larger collecting surface at higher energies.



Water under pressure

A UN analysis sets out global water-management concerns ahead of Earth Summit.

BY NATASHA GILBERT

Water should be at the top of the agenda for the Earth Summit in Rio de Janeiro, Brazil, in June, a United Nations report urges.

The fourth *World Water Development Report* by the UN Educational, Scientific and Cultural Organization (UNESCO), launched at the World Water Forum in Marseilles, France, on 12 March, notes that industry, agriculture and booming urban populations are putting Earth's water supplies under unprecedented pressure (see graphic). Hundreds of millions of people do not have access to clean water, leaving them at risk from waterborne diseases. Without prompt action to improve water-management policies, the report says, a global crisis looms.

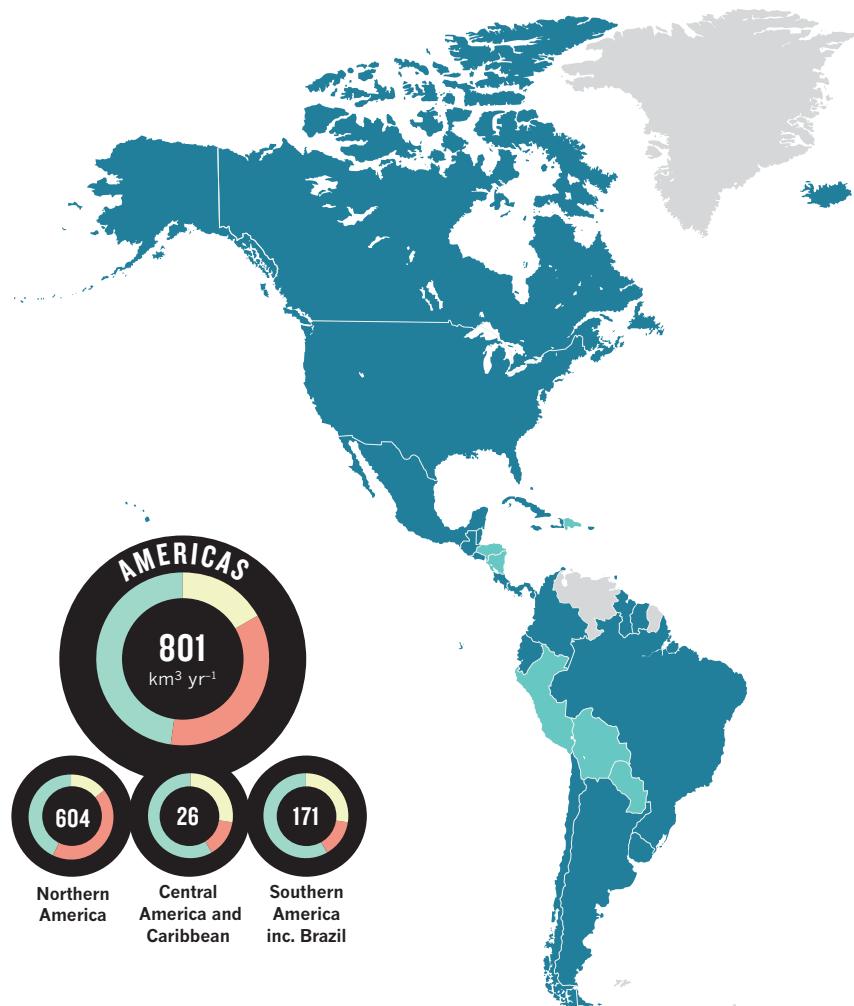
Although the document contains a plethora of facts and figures, its authors argue that a lack of reliable data on water quality and usage has become a stumbling block for efforts to strengthen policies and enforce regulations. "You cannot properly manage something that you don't know about," says Olcay Ünver, coordinator of the UN World Water Assessment Programme. But closing the knowledge gap will be expensive: building a gauging station to measure a river's flow can cost more than US\$1 million, for example, and the expense of ongoing operation can be difficult for poor countries to justify. The report recommends increasing the use of remote-sensing technologies to monitor water quality, but notes that these will never completely replace information gathered on the ground.

The report also focuses on the burgeoning demands of agriculture. Food production already consumes more than two-thirds of the world's extracted water, and food demand is expected to rise by 70% by 2050, owing to population growth. Research into improving crop yields and drought tolerance will help nations to meet needs while using water more efficiently.

The report concludes that policy-makers must balance the requirements of agriculture and industry with the need for sustainable sources of clean drinking water by developing integrated policies that satisfy all three sectors. Michel Jarraud, chairman of UN-Water — a grouping of 28 UN organizations including UNESCO — says that the group will tell leaders at the Rio summit that "the challenges, risks and uncertainties blocking the road to sustainable development require a collective response by the whole international community". ■

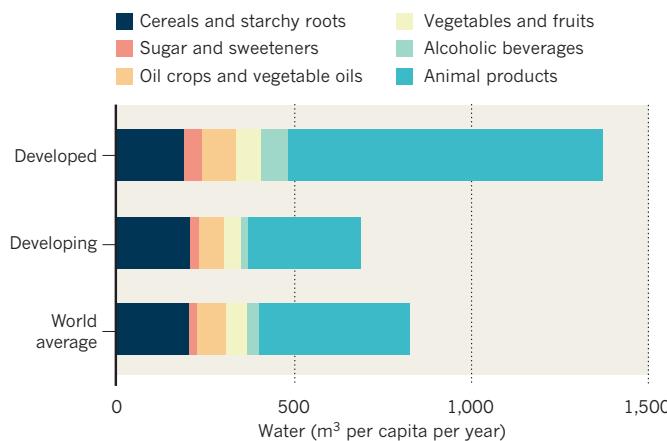
WORLD'S THIRSTIEST

Just three nations — India, China and the United States — together use about one-third of the roughly 4,000 km³ of water extracted globally each year. In general, water demands in developed countries have been declining during the past 20 years, mostly as a result of more efficient use of water resources.



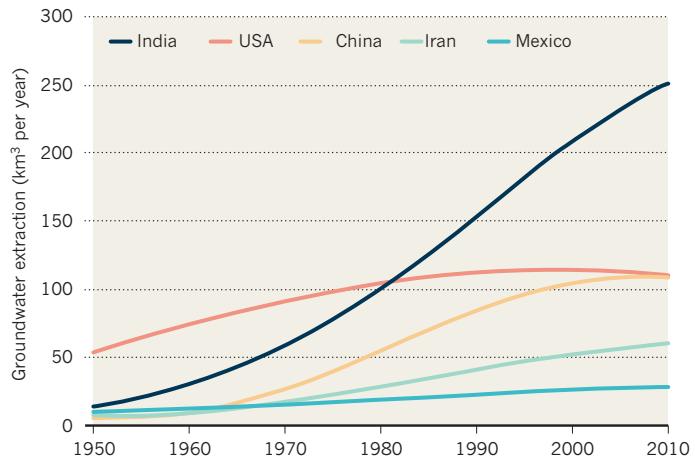
ON THE FARM

Agriculture consumes almost 70% of all extracted water. Animal husbandry is the most water-intensive aspect of farming, and causes the greatest disparity in water consumption between the developed and developing worlds. The United States leads the world in per capita water use attributable to animal products, with each person consuming the equivalent of about 1,200 m³ per year.



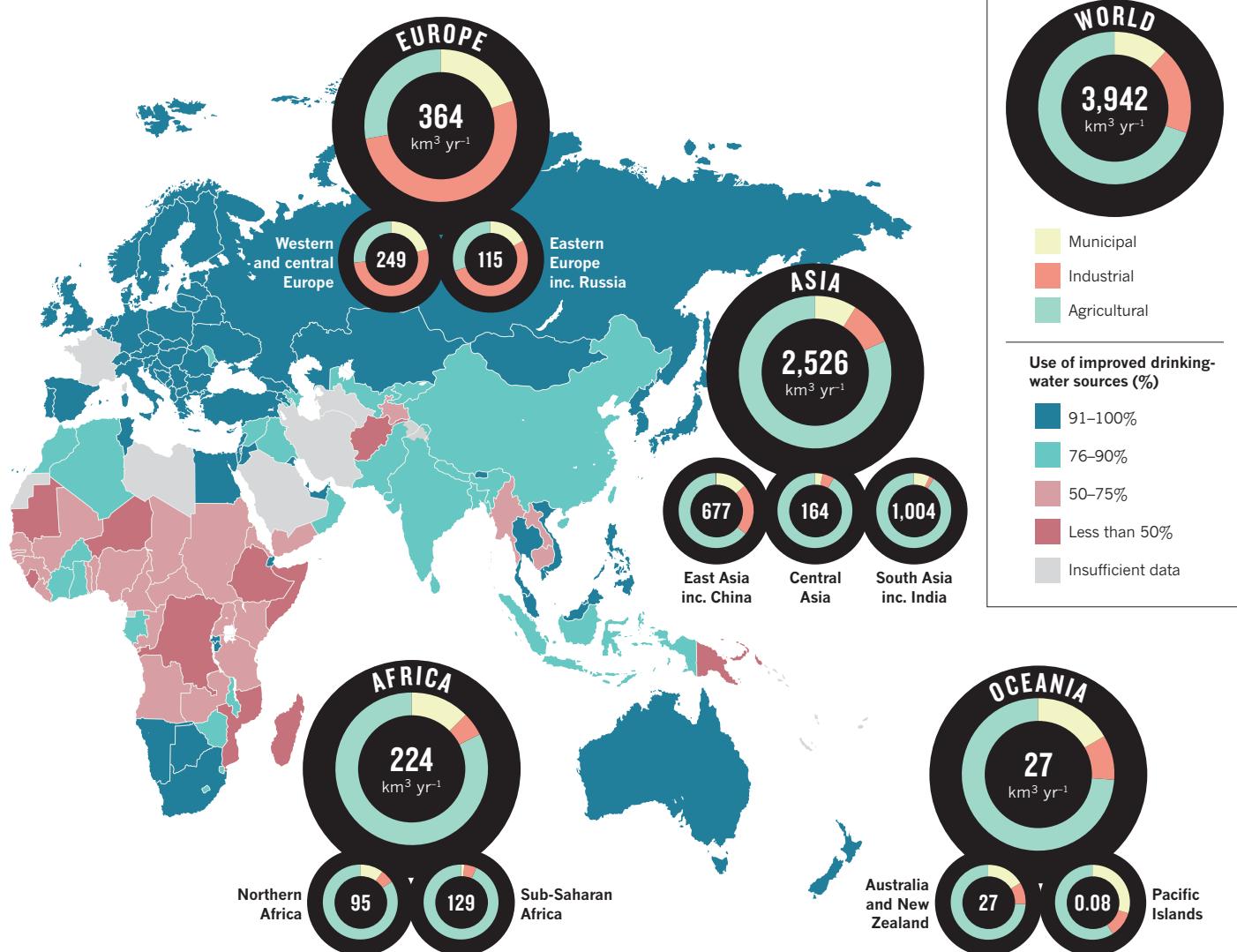
DRILLING DOWN

Countries are increasingly meeting demand by extracting water from non-renewable underground sources. Groundwater extraction has tripled in the past 50 years. India has had the largest growth, jumping from less than 25 km³ in 1950 to 250 km³ in 2010: about one-quarter of the global total (other countries shown for comparison).



DYING FOR A DRINK

Nearly 1 billion people have no access to safe and improved water sources such as boreholes, protected wells, springs and rainwater collections. More than 80% of the world's waste water is not collected or treated, causing millions of deaths from waterborne diarrhoeal diseases every year in the developing world. Urban settlements are the main source of pollution, and the challenge will grow as the world's urban population almost doubles to 6.3 billion by 2050.





China is investing heavily in research and technology to modernize its agriculture.

FUNDING

China's budget backs science

Yet reforms to funding systems and more support for basic research are needed, say scientists.

BY JANE QIU IN BEIJING

The Chinese government's promise last week that researchers will enjoy another year of increased funding was not unexpected, given the country's relatively buoyant economy and high regard for science. But some scientists fear that too little money will be spent on good-quality basic research, and that too much will be wasted in a funding system in which good connections often trump academic expertise.

In a speech at the opening session of the annual National People's Congress, China's Premier, Wen Jiabao, lauded science and technology as key drivers of economic growth and individual prosperity, and backed up the rhetoric with hard cash. This year, central-government expenditure on science and technology is set to rise to 228.5 billion renminbi (US\$36.1 billion), a 12.4% increase on last year's spending, which slightly trails the country's overall projected budget increase of 13.7%. Of that science budget, 32.5 billion renminbi will go towards basic research, a 10.1% increase on the 29.5 billion renminbi spent in 2011. Wen also stressed the government's continuing commitment to improving agriculture, promising

an additional 10.1 billion renminbi — a 53% rise on last year — for developing new agricultural technologies and modernizing China's seed industry. Further details of funding allocations will emerge in the coming months, but sources close to government say that other areas likely to receive increased support include information technology, drug discovery, regenerative medicine, renewable energy and the exploitation of mineral and fuel resources.

In the past decade, China's total expenditure on research and development (R&D) increased by about 20% per year. In 2009, it overtook Japan to become the world's second-largest investor in R&D after the United States. "This increasing flow of investment has seen a jump in the total scientific output," says Cong Cao, a science-policy researcher at the University of Nottingham, UK. Last year, a study by Britain's Royal Society found that, in 2004–08, China produced 10% of the world's published scientific articles, putting it second after the United States.

However, a study conducted by the World Bank and China's cabinet concluded last month that Chinese research quality falls short. It noted that the country

► **NATURE.COM**
Read more at *Nature China*:
nature.com/nchina

produces relatively few high-impact articles, and that the majority of Chinese patents constitute minor novelties rather than genuine innovations. These problems are "less about the amount of money spent in science than how the money is spent", says Su Jun, a science-policy researcher at Tsinghua University in Beijing.

BACK TO BASICS

Experts say that one factor holding back China's research quality is its modest spending on basic research. According to the annual report released last month by the National Bureau of Statistics of China, the country's total public and private R&D expenditure last year was 861 billion renminbi (1.83% of gross domestic product), of which only 39.6 billion renminbi (4.6%) was spent on basic research. This falls far behind the 15–25% share in most developed countries.

"Weak capacity in basic science has become the main constraining factor for most R&D in China, including drug discovery," says Liu Gang, a pharmacologist at Tsinghua University. In 2006, China unveiled an ambitious plan to invigorate its drug industry, with funding of about 20 billion renminbi over 15 years. "I'm not optimistic that we will be able to get any blockbuster drug out of that programme," says Liu.

Some researchers also criticize the large share of funding typically given to government-led initiatives. "Genuine scientific breakthroughs cannot rely on such top-down mega-projects," says Mu-Ming Poo, director of the Chinese Academy of Sciences' Institute of Neuroscience in Shanghai. The United States, by contrast, funnels a large proportion of spending to investigator-initiated research. In China, such research is mostly sponsored by the National Natural Science Foundation, but its funding, which amounted to 12.4 billion renminbi last year, is only a small proportion of the government's science expenditure.

Lack of fairness and transparency in funding decisions — with success often depending on personal connections — could also be holding back Chinese science. "A huge amount of money is often controlled by a small number of people, and there are significant irregularities in their funding decisions," says one Beijing-based researcher. And in comparison with other research expenditure, investment in human capital is limited, with many young scientists deterred by poor salaries, adds Su. "Most labs in China struggle to find good postdocs, the key workforce of research," agrees Poo.

The science ministry is under growing pressure from the scientific community to reform the research funding and appraisal system to help address these problems. Many researchers contacted by *Nature* hope that China will make the changes a top priority after its new government takes office later this year. "The extent to which the reform can be instigated and successful will determine the future of China's technological and economic competitiveness," says Cao. ■

ETHICS

Japan fails to settle university dispute

Investigations highlight need for a national, independent body to oversee research ethics.

BY DAVID CYRANOSKI IN TOKYO

It has been a rough year for materials scientist Akihisa Inoue, the president of Tohoku University in Japan.

Last March, an earthquake crippled his campus (see *Nature* **483**, 141–143; 2012). Since then, he has had to retract a series of papers because they contained text that had appeared in his previous publications, and has faced continuing calls for his resignation from the university, which he has rejected. His critics, mostly professors at his university, claim that some of his work cannot be replicated, and that there are irregularities in the data in some of his papers (see *Nature* **470**, 446–447; 2011).

Inoue denies any manipulation of data, and there is no evidence that he has committed any scientific misconduct. Indeed, with more than 2,500 publications to his name, Inoue is one of the world's leading experts in metallic glasses, materials that are more elastic and more resistant to corrosion than metals. He has previously told *Nature* that other researchers may simply lack the skills and experience to reproduce some of his lab's results.

Yet Inoue's battle with his detractors is far from over. Since January, two inquiries into his research have reported their conclusions. One offered a rebuke for the duplications; the other recommended a further investigation, giving his critics renewed vigour. The row raises questions about how universities in Japan should investigate allegations against their most senior staff, given that the country has no external body with this responsibility.

In December 2007, a committee from Tohoku University dismissed the need for an official investigation into Inoue, on the basis that there was no case against him. But Inoue's critics were not appeased, because, they argued, the majority of the committee had been appointed to their administrative posts (but not to the committee itself) by Inoue. The committee denied any conflict of interest, adding that it was able to judge the case fairly.

Since then, materials scientist Fumio Saito at Tohoku has pointed out that the text in seven of Inoue's papers substantially duplicated work previously published by Inoue's lab. These seven papers have since been retracted. Inoue told *Nature* that the duplications were

accidents, or the result of miscommunication with co-authors.

Under pressure to take a closer look at Inoue's work, Tohoku's directors assembled another committee in February 2011. On 24 January, that committee's final report concluded that although Inoue was at fault for the duplications, and that it should not happen again, the reuse of text is, to some degree, accepted practice in materials science, particularly in papers that draw on conference proceedings.

The Japan Science and Technology Agency (JST), a government funding body that gave Inoue ¥2.1 billion (US\$26 million at the current exchange rate) to support his research from 1997 to 2007, also launched an investigation. It commissioned Makoto Misono, head of the Japan Union of Chemical Science and Technology, to lead a committee to investigate whether the duplications and alleged data irregularities called any of Inoue's main results into question. Its report, released on 17 February, concluded that they had not.

The committee did not investigate any other allegations, but suggested that to settle the matter once and for all, an independent body should conduct an inquiry. The university says it is already conducting the investigation requested by the JST, but has declined to provide any details about its remit. Inoue has declined to comment on Misono's report, or on its call for an independent investigation.

The broader message of the affair, Misono says, is that Japan should give more responsibility for the oversight of research conduct to an independent body such as the Science Council of Japan, an advisory body to the government. In general, Misono says, "when evaluating claims of scientific misconduct, it is necessary to have objective and fair investigation. If a research institution cannot do that, it's necessary to have an independent third party." ■ **SEE EDITORIAL, P.246**

CORRECTION

The News story 'Trouble at the text mine' (*Nature* **483**, 134–135; 2012) wrongly located BioNOT at the University of Wisconsin-Madison instead of the University of Wisconsin-Milwaukee.

A TALE OF

1

Since the 1960s, researchers have been scrutinizing a handful of patients who underwent a radical kind of brain surgery. The cohort has been a boon to neuroscience — but soon it will be gone.

TWO HALVES

BY DAVID WOLMAN

In the first months after her surgery, shopping for groceries was infuriating. Standing in the supermarket aisle, Vicki would look at an item on the shelf and know that she wanted to place it in her trolley — but she couldn't. “I'd reach with my right for the thing I wanted, but the left would come in and they'd kind of fight,” she says. “Almost like repelling magnets.” Picking out food for the week was a two-, sometimes three-hour ordeal. Getting dressed posed a similar challenge: Vicki couldn't reconcile what she wanted to put on with what her hands were doing. Sometimes she ended up wearing three outfits at once. “I'd have to dump all the clothes on the bed, catch my breath and start again.”

In one crucial way, however, Vicki was better than her pre-surgery self. She was no longer racked by epileptic seizures that were so severe they had made her life close to unbearable. She once collapsed onto the bar of an old-fashioned oven, burning and scarring her back. “I really just couldn't function,” she says. When, in 1978, her neurologist told her about a radical but dangerous surgery that might help, she barely hesitated. If the worst were to happen, she knew that her parents would take care of her young daughter. “But of course I worried,” she says. “When you get your brain split, it doesn't grow back together.”

In June 1979, in a procedure that lasted nearly 10 hours, doctors created a firebreak to contain Vicki's seizures by slicing through her corpus callosum, the bundle of neuronal fibres connecting the two sides of her brain. This drastic procedure, called a corpus callosotomy, disconnects the two sides of the neocortex, the home of language, conscious thought and movement control. Vicki's supermarket predicament was

the consequence of a brain that behaved in some ways as if it were two separate minds.

After about a year, Vicki's difficulties abated. “I could get things together,” she says. For the most part she was herself: slicing vegetables, tying her shoe laces, playing cards, even waterskiing.

But what Vicki could never have known was that her surgery would turn her into an accidental superstar of neuroscience. She is one of fewer than a dozen ‘split-brain’ patients, whose brains and behaviours have been subject to countless hours of experiments, hundreds of scientific papers, and references in just about every psychology textbook of the past generation. And now their numbers are dwindling.

Through studies of this group, neuroscientists now know that the healthy brain can look like two markedly different machines, cabled together and exchanging a torrent of data. But when the primary cable is severed, information — a word, an object, a picture — presented to one hemisphere goes unnoticed in the other. Michael Gazzaniga, a cognitive neuroscientist at the University of California, Santa Barbara, and the godfather of modern split-brain science, says that even after working with these patients for five decades, he still finds it thrilling to observe the disconnection effects first-hand. “You see a split-brain

ONATURE.COM
To hear more about split-brain patients, visit: go.nature.com/knhmxk

patient just doing a standard thing — you show him an image and he can't say what it is. But he can pull that same object out of a grab-bag,” Gazzaniga says. “Your heart just races!”

Work with the patients has teased out differences between the two hemispheres,

revealing, for instance, that the left side usually leads the way for speech and language computation, and the right specializes in visual-spatial processing and facial recognition. "The split work really showed that the two hemispheres are both very competent at most things, but provide us with two different snapshots of the world," says Richard Ivry, director of the Institute of Cognitive and Brain Sciences at the University of California, Berkeley. The idea of dichotomous consciousness captivated the public, and was greatly exaggerated in the notion of the 'creative right brain'. But further testing with split-brain patients gave a more-nuanced picture. The brain isn't like a computer, with specific sections of hardware charged with specific tasks. It's more like a network of computers connected by very big, busy broadband cables. The connectivity between active brain regions is turning out to be just as important, if not more so, than the operation of the distinct parts. "With split-brain patients, you can see the impact of disconnecting a huge portion of that network, but without damage to any particular modules," says Michael Miller, a psychologist at the University of California, Santa Barbara.

David Roberts, head of neurosurgery at Dartmouth-Hitchcock Medical Center in Lebanon, New Hampshire, sees an important lesson in split-brain research. He operated on some of the cohort members, and has worked closely with Gazzaniga. "In medical school, and science in general, there is so much emphasis on large numbers, labs, diagnostics and statistical significance," Roberts says — all crucial when, say, evaluating a new drug. But the split-brain cohort brought home to him how much can be gleaned from a single case. "I came to learn that one individual, studied well, and thoughtfully, might enable you to draw conclusions that apply to the entire human species," he says.

Today, the split-brain patients are getting on in years; a few have died, one has had a stroke and age in general has made them all less fit for what can be taxing research sessions of sitting, staring and concentrating. The surgery, already quite rare, has been replaced by drug treatments and less drastic surgical procedures. Meanwhile, imaging technologies have become the preferred way to look at brain function, as scientists can simply watch which areas of the brain are active during a task.

But to Miller, Ivry, Gazzaniga and others, split-brain patients remain an invaluable resource. Imaging tools can confirm, for example, that the left hemisphere is more active than the right when processing language. But this is dramatically embodied in a split-brain patient, who may not be able to read aloud a word such as 'pan' when it's presented to the right hemisphere, but can point to the appropriate drawing. "That gives you a sense of the right hemisphere's ability to read, even if it can't access the motor system to produce speech," Ivry says. "Imaging is very good for telling you where something happens," he adds, "whereas patient work can tell you how something happens."

A CABLE, CUT

Severing the corpus callosum was first used as a treatment for severe epilepsy in the 1940s, on a group of 26 people in Rochester, New York. The aim was to limit the electrical storm of the seizure to one side of the brain. At first, it didn't seem to work. But in 1962, one patient showed significant improvement. Although the procedure never became a favoured treatment strategy — it's invasive, risky, and drugs can ease symptoms in many people — in the decades since it nevertheless became a technique of last resort for treating intractable epilepsy.

To Roger Sperry, then a neurobiologist and neuropsychologist at the California Institute of Technology, and Gazzaniga, a graduate student in Sperry's lab, split-brain patients presented a unique opportunity to explore the lateralized nature of the human brain. At the time, opinion on the matter was itself divided. Researchers who studied the first split-brain patients in the 1940s had concluded that the separation didn't noticeably affect thought or behaviour. (Gazzaniga and others suspect that these early sections were incomplete, which might also explain why they didn't help the seizures.) Conversely, studies conducted by Sperry and colleagues in the 1950s revealed greatly altered

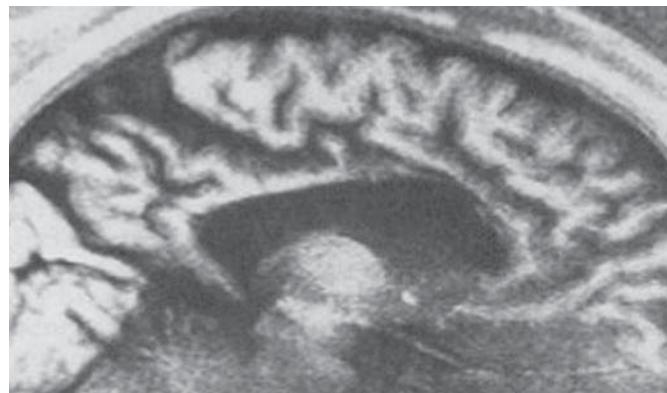
brain function in animals that had undergone callosal sections. Sperry and Gazzaniga became obsessed with this inconsistency, and saw in the split-brain patients a way to find answers.

The duo's first patient was a man known as W.J., a former Second World War paratrooper who had started having seizures after a German soldier clocked him in the head with the butt of a rifle. In 1962, after W.J.'s operation, Gazzaniga ran an experiment in which he asked W.J. to press a button whenever he saw an image. Researchers would then flash images of letters, light bursts and other stimuli to his left or right field of view. Because the left field of view is processed by the right hemisphere and vice versa, flashing images quickly to one side or the other delivers the information solely to the intended hemisphere (see 'Of two minds').

For stimuli delivered to the left hemisphere, W.J. showed no hang-ups; he simply pressed the button and told the scientists what he saw. With the right hemisphere, W.J. said he saw nothing, yet his left hand kept pressing the button every time an image appeared. "The left and right didn't know what the other was doing," says Gazzaniga. It was a paradigm-blasting discovery showing that the brain is more divided than anyone had predicted¹.

Suddenly, the race was on to delve into the world of lateralized function. But finding more patients to study proved difficult. Gazzaniga estimates that at least 100 patients, and possibly many more, received a corpus callosotomy. But individuals considered for the operation tend to have other significant developmental or cognitive problems; only a few have super-clean cuts and are neurologically healthy enough to be useful to researchers. For a while, Sperry, Gazzaniga and their colleagues didn't know if there was ever going to be anyone else like W.J..

But after contacting neurosurgeons, partnering with epilepsy centres and assessing many potential patients, they were able to identify a few suitable people in California, then a cluster from the eastern part of the United States, including Vicki. Through the 1970s and the early

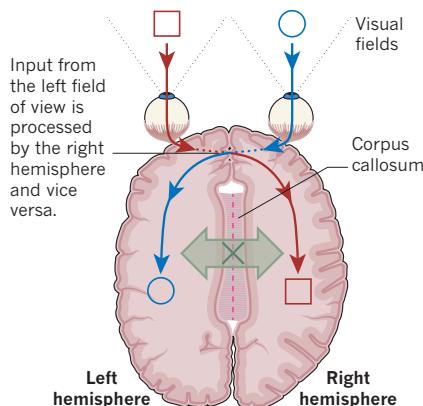


The corpus callosum (bright white in top image) connects the two hemispheres of the brain. After a corpus callosotomy, the callosum tissue retracts (bottom image), leaving just the ventricle (black).

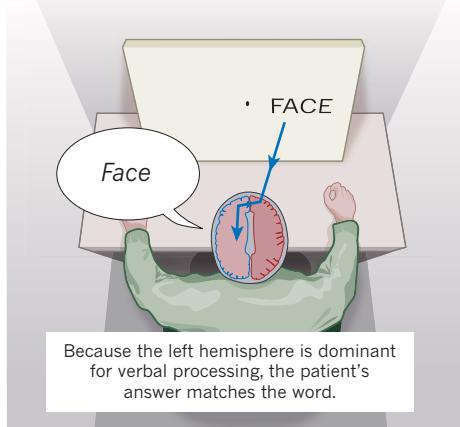
OF TWO MINDS

Experiments with split-brain patients have helped to illuminate the lateralized nature of brain function.

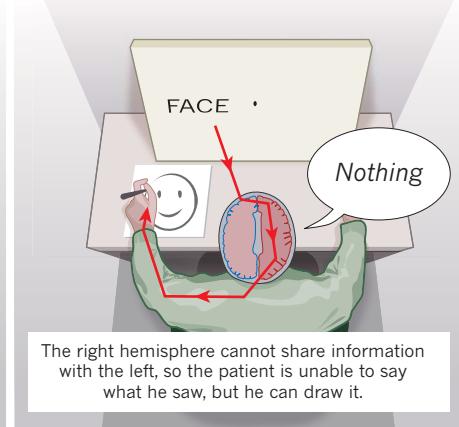
Split-brain patients have undergone surgery to cut the corpus callosum, the main bundle of neuronal fibres connecting the two sides of the brain.



A word is flashed briefly to the right field of view, and the patient is asked what he saw.



Now a word is flashed to the left field of view, and the patient is asked what he saw.



1980s, split-brain research expanded, and neuroscientists became particularly interested in the capabilities of the right hemisphere — the one conventionally believed to be incapable of processing language and producing speech.

Gazzaniga can tick through the names of his “endlessly patient patients” with the ease of a proud grandparent doing a roll call of grandchildren — W.J., A.A., R.Y., L.B., N.G.. For medical confidentiality, they are known in the literature by initials only. (Vicki agreed to be identified in this article, provided that her last name and hometown were not published.)

On stage last May, delivering a keynote address at the Society of Neurological Surgeons’ annual meeting in Portland, Oregon, Gazzaniga showed a few grainy film clips from a 1976 experiment with patient P.S., who was only 13 or 14 at the time. The scientists wanted to see his response if only his right hemisphere saw written words.

In Gazzaniga’s video, the boy is asked: who is your favourite girlfriend, with the word girlfriend flashed only to the right hemisphere.

callosotomy, it turned out, resulted in some universal disconnections, but also affected individuals very differently.

In 1981, Sperry was awarded a share of the Nobel Prize in Physiology or Medicine for the split-brain discoveries. (“He deserved it,” Gazzaniga says.) Sperry died in 1994, but by that point, Gazzaniga was leading the charge. By the turn of the century, he and other split-brain investigators had turned their attention to another mystery: despite the dramatic effects of callosotomy, W.J. and later patients never reported feeling anything less than whole. As Gazzaniga wrote many times: the hemispheres didn’t miss each other.

Gazzaniga developed what he calls the interpreter theory to explain why people — including split-brain patients — have a unified sense of self and mental life³. It grew out of tasks in which he asked a split-brain person to explain in words, which uses the left hemisphere, an action that had been directed to and carried out only by the right one. “The left hemisphere made up a post hoc answer that fit the situation.” In one of Gazzaniga’s favourite examples, he flashed the word ‘smile’ to a patient’s right hemisphere and the word ‘face’ to the left hemisphere, and asked the patient to draw what he’d seen. “His right hand drew a smiling face,” Gazzaniga recalled. “Why did you do that?” I asked. He said, ‘What do you want, a sad face? Who wants a sad face around?’ The left-brain interpreter, Gazzaniga says, is what everyone uses to seek explanations for events, triage the barrage of incoming information and construct narratives that help to make sense of the world.

The split-brain studies constitute “an incredible body of work”, said Robert Breeze, a neurosurgeon at the University of Colorado Hospital in Aurora, after listening to Gazzaniga’s lecture last year. But Breeze, like many other neuroscientists, sees split-brain research as outdated. “Now we have technologies that enable us to see these things” — tools such as functional magnetic resonance imaging (fMRI) that show the whereabouts of brain function in great detail.

Miller, however, disagrees. “These kinds of patients can tell us things that fMRI can never tell us,” he says.

SUBJECT OF INTEREST

Seated at a small, oval dining-room table, Vicki faces a laptop propped up on a stand, and a console with a few large red and green buttons. David Turk, a psychologist at the University of Aberdeen, UK, has flown in for the week to run a series of experiments.

Vicki’s grey-white hair is pulled back in a ponytail. She wears simple white sneakers and, despite the autumn chill, shorts. She doesn’t want to get too warm: when that happens she can get drowsy and lose focus, which can wreck a whole day of research.

I HAVE A HARD TIME SAYING IT'S ALL OVER.

As predicted, the boy can’t respond verbally. He shrugs and shakes his head, indicating that he doesn’t see any word, as had been the case with W.J.. But then he giggles. It’s one of those tell-tale teen giggles — a soundtrack to a blush. His right hemisphere has seen the message, but the verbal left-hemisphere remains unaware. Then, using his left hand, the boy slowly selects three Scrabble tiles from the assortment in front of him. He lines them up to spell L-I-Z: the name, we can safely assume, of the cute girl in his class. “That told us that he was capable of language comprehension in the right hemisphere,” Gazzaniga later told me. “He was one of the first confirmation cases that you could get bilateral language — he could answer queries using language from either side.”

The implications of these early observations were “huge”, says Miller. They showed that “the right hemisphere is experiencing its own aspect of the world that it can no longer express, except through gestures and control of the left hand”. A few years later, the researchers found that Vicki also had a right-hemisphere capacity for speech². Full

During a break, Vicki fetches an old photo album. In one picture, taken soon after her surgery, she is sitting up in the hospital bed. Her hair is starting to grow back as black stubble and she and her daughter have wide smiles. Another page of the album has a slightly faded printout of a 1981 paper from *The Journal of Neuroscience* glued into it: the first published report involving data gleaned from Vicki, in which researchers describe how she, like PS., had some capacity for language in her right hemisphere⁴.

When pressed to share the most difficult aspect of her life in science, the perpetually upbeat Vicki says that it would have to be an apparatus called the dual Purkinje eye tracker. This medieval-looking device requires the wearer to bite down on a bar to help keep the head still so that researchers can present an image to just the left or right field of view. It is quite possible that Vicki has spent more of her waking hours biting down on one of those bars than anyone else on the planet.

Soon, it is time to get back to work. Turk uses some two-sided tape to affix a pair of three-dimensional glasses onto the front of Vicki's thin, gold-rimmed bifocals. The experiment he is running aims to separate the role of the corpus callosum in visual processing from that of deeper, 'subcortical' connections unaffected by the callosotomy. Focusing on the centre of the screen, Vicki is told to watch as the picture slowly switches between a house and different faces — and to press the button every time she sees the image change. Adjusting her seat, she looks down the bridge of her nose at the screen and tells Turk that she's ready to begin.

DEEP CONNECTIONS

Other researchers are studying the role of subcortical communication in the coordinated movements of the hands. Split-brain patients have little difficulty with 'bimanual' tasks, and Vicki and at least one other patient are able to drive a car. In 2000, a team led by Liz Franz at the University of Otago in New Zealand asked split-brain patients

to carry out both familiar and new bimanual tasks. A patient who was an experienced fisherman, they found, could pantomime tying a fishing line, but not the unfamiliar task of threading a needle. Franz concluded that well-practised bimanual skills are coordinated at the subcortical level, so split-brain people are able to smoothly choreograph both hands⁵.

Miller and Gazzaniga have also started to study the right hemisphere's role in moral reasoning. It is the kind of higher-level function for which the left hemisphere was assumed to be king. But in the past few years, imaging studies have shown that the right hemisphere is heavily involved in the processing of others' emotions, intentions and beliefs — what many scientists have come to understand as the 'theory of mind'⁶. To Miller, the field of enquiry perfectly illustrates the value of split-brain studies because answers can't be found by way of imaging tools alone.

In work that began in 2009, the researchers presented two split-brain patients with a series of stories, each of which involved either accidental or intentional harm. The aim was to find out whether the patients felt that someone who intends to poison his boss but fails because he mistakes sugar for rat poison, is on equal moral ground with someone who accidentally kills his boss by mistaking rat poison for sugar⁷. (Most people conclude that the former is more morally reprehensible.) The researchers read the stories aloud, which meant that the input was directed to the left hemisphere, and asked for verbal responses, so that the left hemisphere, guided by the interpreter mechanism, would also create and deliver the response. So could the split-brain patients make a conventional moral judgement using just that side of the brain?

No. The patients reasoned that both scenarios were morally equal. The results suggest that both sides of the cortex are necessary for this type of reasoning task.

But this finding presents an additional puzzle, because relatives and friends of split-brain patients do not notice unusual reasoning or theory-of-mind deficits. Miller's team speculates that, in everyday life, other reasoning mechanisms may compensate for disconnection effects that are exposed in the lab. It's an idea that he plans to test in the future.

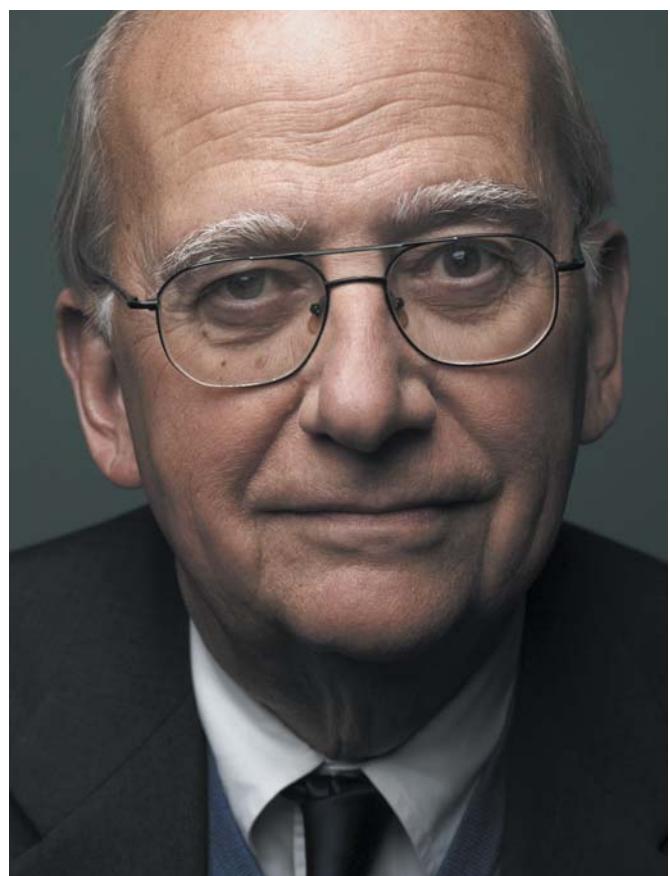
As the opportunities for split-brain research dwindle, Gazzaniga is busy trying to digitize the archive of recordings of tests with cohort members, some of which date back more than 50 years. "Each scene is so easy to remember for me, and so moving," he says. "We were observing so many astonishing things, and others should have the same opportunity through these videos." Perhaps, he says, other researchers will even uncover something new.

Other split-brain patients may become available — there is a small cluster in Italy, for instance. But with competition from imaging research and many of the biggest discoveries about the split brain behind him, Gazzaniga admits that the glory days of this field of science are probably gone. "It is winding down in terms of patients commonly tested." Still, he adds: "I have a hard time saying it's all over."

And maybe it's not — as long as there are scientists pushing to tackle new questions about lateralized brain function, connectivity and communication, and as long as Vicki and her fellow cohort members are still around and still willing participants in science. Her involvement over the years, Vicki says, was never really about her. "It was always about getting information from me that might help others." ■

David Wolman is a freelance writer based in Portland, Oregon, and the author, most recently, of *The End of Money*.

1. Gazzaniga, M. S., Bogen, J. E. & Sperry, R. W. *Proc. Natl Acad. Sci. USA* **48**, 1765–1769 (1962).
2. Gazzaniga, M. S. *Brain* **123**, 1293–1326 (2000).
3. Gazzaniga, M. S. *Science* **245**, 947–952 (1989).
4. Sidtis, J. J., Volpe, B. T., Wilson, D. H., Rayport, M. & Gazzaniga, M. S. *J. Neurosci.* **1**, 323–331 (1981).
5. Franz, E. A., Walidie, K. E. & Smith, M. J. *Psychol. Sci.* **11**, 82–85 (2000).
6. Young, L. & Saxe, R. *NeuroImage* **40**, 1912–1920 (2008).
7. Miller, M. B. *et al. Neuropsychologia* **48**, 2215–2220 (2010).



Michael Gazzaniga has worked with split-brain patients for 50 years.



Make it a decaf

The enduring quest for a coffee bean without the buzz.

BY BRENDAN BORRELL

P

Paulo Mazzafera punched a pea-sized disc out of a waxy green coffee leaf, then placed the disc in a small vial with a mixture of chloroform and methanol to dissolve it. Later, he loaded the extract, along with 95 other samples, into a high-performance liquid chromatography machine, which separates out each chemical component. When the plant physiologist returned to his lab at the University of Campinas in Brazil the next morning, he sat down at his laptop to examine the results. Scrolling from one chromatogram to the next, he scrutinized the peak representing caffeine. In one plant, it was missing.

IMAGE SOURCE/CORBIS

Mazzafera ran the sample twice more and then, just before noon, called his collaborator Bernadete Silvarolla, based at the agricultural station nearby, to share the news. "Are you sure?" she asked. He was. In fact, he was thrilled. After screening thousands of plants over the course of two decades, his project to find a naturally caffeine-free coffee finally seemed to be bearing fruit. That was in late 2003.

Coffee contains some 2,000 chemical compounds that give the drink its enticing aroma and flavour, including caffeine, a stimulant and natural pesticide. Removing the caffeine while leaving all the others intact poses a significant challenge. Brewers have generally turned to chemistry: Ludwig Roselius of Bremen, Germany, patented the first commercial decaffeination process in 1905. But his coffee, marketed as Kaffee HAG, used benzene in the extraction process, and the chemical was later replaced by less toxic solvents. Today, companies may instead douse raw green coffee beans in high-pressure liquid carbon dioxide or soak them in hot water for several hours to remove the caffeine before roasting. Aficionados say that all these methods destroy the taste, but the decaf market is still worth US\$2 billion a year.

Researchers have long sought a better bean, harvested directly from the plant caffeine-free. This would preserve coffee's complex flavour and give growers a high-end slice of the decaf market. But developing such a bean through conventional breeding or even genetic modification has proved more difficult than anyone

NATURE.COM

For pictures from the field and an interview with the author, see: go.nature.com/qvscwj

anticipated (see 'The ups and downs of decaffeination'). Coffee plants take years to begin producing beans, and can be fickle when they do. Moreover, to make them profitable to farm, the plants need to be productive, ripen synchronously and be of a size and shape that can be harvested easily by hand or by machines. The loss of any of these traits can render a plant worthless. The caffeine-free quest has produced a string of high-profile papers, but not a drop of marketable coffee.

Basic research often proves hard to translate into industrial agriculture, says Rod Sharp, a retired plant cell biologist who tried to produce caffeine-free coffee in the 1980s while at DNA Plant Technology in Cinnaminson, New Jersey. "Those in the research lab are always a bit naive. We jump up and down when there is a breakthrough, but turning that into a commercial operation is another challenge altogether." However, Sharp remains confident that the plant will one day yield its secrets. "It will happen," he says. "It's just a long incubation period." This is certainly true for Mazzafera's plant: more than eight years after its discovery, his colleagues are still trying to turn it into a crop.

Coffee is worth a total of \$15 billion to \$20 billion per year to exporting countries, which include Brazil, Colombia and Vietnam. A relatively recent innovation, compared with tea or wine, it dates to about the fifteenth century when — according to at least one account — a Yemeni mystic described a revitalizing beverage prepared in Ethiopia by roasting and boiling berries. Two coffee species dominate the market today: *Coffea arabica*, the better-tasting bean, which grows in cooler climates, and *Coffea canephora*, commonly known as 'robusta', which is used mainly in instant coffee and lower-quality blends. Caffeine, which is present at 1.2% in commercial *C. arabica* and 2–3% in *C. canephora*, has helped to make both species part of a worldwide addiction.

But the stimulant beloved by many is avoided by others who are unusually sensitive to its effects, abstain for religious reasons or simply don't want to be kept awake. Gabriel Bertrand of the Pasteur Institute in Paris discovered a caffeine-free species of coffee on Grande Comore island near Madagascar in 1901. In fact, many of the hundred or so species of *Coffea* are either caffeine-free or contain low levels of the stimulant. Several naturally occurring intermediate-caffeine coffees (with 0.6–1% caffeine) have already reached the market, including one produced by the Italian coffee-maker Illy. Unfortunately, most plants with the lowest caffeine levels produce few beans, or contain high concentrations of bitter compounds. Nevertheless, the existence of such natural variation suggests that a diligent breeder could create a commercially viable caffeine-free strain.

COFFEE AND DREAMS

Mazzafera began his attempts to do just that in 1983 at the Agronomical Institute of Campinas (IAC), a century-old agricultural station in the rolling hills northwest of São Paulo, Brazil. He set out to study the genetics and physiology of caffeine biosynthesis under Alcides Carvalho, a pioneering plant breeder who established the IAC's collection of coffee plants, which now number 70,000 and represent more than 1,000 wild strains, breeding lines, hybrids, mutants and cultivated varieties from around the world.

At first, Mazzafera used an old-school spectrophotometer to measure caffeine content one sample at a time. In 1987, he got a job at the University of Campinas and installed a high-performance liquid chromatography machine in his lab, allowing him to process samples more efficiently. By this point, scientists had sketched out the basic four-step pathway by which *C. arabica* synthesizes caffeine. Mazzafera studied caffeine production and breakdown in great detail in seven species, hoping to find one with defects in a pathway that rendered the plant low in caffeine. At the same time, he and Carvalho, who died in 1993, were cross-breeding commercial coffee cultivars with wild non-*arabica* species low in caffeine. But it proved impossible to eliminate caffeine while maintaining the desirable attributes of *C. arabica*. "We

were just wasting time," says Mazzafera.

In 2000, Mazzafera teamed up with Silvarolla, a coffee breeder at the IAC. They shifted their focus to a group of *C. arabica* plants originally collected during a 1964 United Nations expedition to Eritrea and Ethiopia. Seed samples — 620 in total — were divided up and grown in several countries, including Costa Rica. Later, 308 of these lineages were collected in Costa Rica and sent to Brazil. Mazzafera believed it would be much easier to produce marketable coffee by starting with the Ethiopian *C. arabica* plants than by hybridizing with other species.

It was from this collection that Mazzafera discovered the promising strain in 2003, as well as two more like it. After confirming that, like the leaves, the beans were caffeine-free, he worked out that the plants were defective in the final step of the chemical pathway that turns theobromine — a mild stimulant and diuretic — into caffeine¹. The Brazilian government offered the research group a \$1.2-million grant along with an order to keep the location of the precious plants under wraps. Mazzafera felt certain that commercial growers would be planting the new variety in five years. That is, so long as others didn't get there first.

BEAN MODIFICATION

The advent of genetic engineering led many scientists to try to make decaf by splicing the right genes into the right beans. But coffee has proved resistant to this kind of tinkering. In 1992, geneticist John Stiles of the University of Hawaii in Honolulu wanted to use 'antisense' technology, whereby a gene inserted into the plant reduces production of a target protein. This was the technology used to produce the Flavr Savr tomato, the first genetically modified organism to be approved for human consumption. Stiles's goal was to target a protein in the caffeine-producing pathway.

Problems cropped up almost immediately. Creating any genetically modified plant involves culturing a blob of plant cells on nutrient-rich agar, inserting the desired genetic material into those cells, and then convincing the tissue to sprout into a plant. For *C. arabica*, that process is mysteriously inefficient.

Over the next seven years, Stiles and two postdocs, Kabi Neupane and Stefan Moisyadi, struggled to overcome the biological roadblocks.

They produced plants that seemed to have low levels of caffeine, and Stiles made some enthusiastic claims. In August 1999, for example, he told *The Wall Street Journal* that he would be beginning field trials that month in Hawaii before expanding to Mexico, with commercial prospects after three years.

The plants, however, did not cooperate. As they grew, their caffeine levels rose. Moisyadi and Neupane moved on to academic careers and, in 2000, Stiles left the university, setting up a private lab in Waialua. In 2008, after disagreements with the local community and the state legislature over the right to field test his transgenic coffee, the company folded, and Stiles now says that he was never 100% certain that he had created caffeine-free coffee. Many coffee researchers also doubt that he had. "We were always doing this on a shoestring," Stiles says. "We were never Monsanto."

Neither was Shinjiro Ogita, a postdoc in Hiroshi Sano's lab at the Nara Institute of Science and Technology in Japan. In 2001, he began a research programme targeting an enzyme in the caffeine pathway that had recently been identified in tea. His group used an efficient gene-silencing technology called RNA interference, and worked with robusta coffee in the hope that cell culture would be easier. It wasn't. Few cells took up the engineered DNA. Ogita was able to recover just enough to produce 35 transgenic seedlings.

He tested some leaves and found they had as much as 70% less caffeine than his control plants. "It was kind of unbelievable," he says. He remembers popping a bottle of Dom Perignon champagne the day

The ups and downs of decaffeination

For more than a century, brewers and growers have looked for ways to create tasty, caffeine-free coffee.



Chemistry

▲ Several well-established chemical methods strip beans of their caffeine, which can be sold for use in drugs and cosmetics.

▼ Extraction processes can be very expensive. Coffee aficionados say that the flavour is destroyed.



Hybridization

▲ Caffeine-free species of *Coffea* might be crossed with better-tasting *Coffea arabica*.

▼ Maintaining all desired traits is complex. One potential low-caffeine (~0.37%) strain is a hybrid of three species.



Breeding

▲ Maintaining the desired traits of *Coffea arabica* might be easier if it were bred with a caffeine-free strain of the same species.

▼ Caffeine-free strains are hard to find. Some undesirable traits may still be difficult to 'breed out'.



Genetic engineering

▲ Can specifically target caffeine-biosynthesis genes.

▼ Coffee species are notoriously difficult to engineer. Public acceptance of genetically modified foods is low.



Mutagenization

▲ A well-accepted method for creating variation in crop traits.

▼ Resultant specimens need to be screened, which can be a long and arduous process. Non-targeted traits may be affected.

his paper was accepted by *Nature*². He has since tailored the methods to *C. arabica*, but the plants have still not produced beans. Now at Toyama Prefectural University, Ogita tends about 40 transgenic plants. Each year, he says, the female part of the flowers, called the pistil, matures and deteriorates a week before the pollen is ready.

Even if Ogita can overcome these breeding issues, it is hard to imagine that transgenic caffeine-free coffee will land on supermarket shelves any time soon. One basic hurdle, says biochemist Alan Crozier at the University of Glasgow, UK, is that side routes in the caffeine pathway result in some production of caffeine, so transgenic coffee may never be caffeine-free. Gaining public acceptance is likely to be another challenge. Monsanto and other agricultural giants have generally commercialized technologies with the grower in mind — for example, pest resistance or herbicide tolerance — rather than focus on consumer-oriented traits such as low caffeine or high antioxidants. So finding financial backing may be tricky.

BREED APPEAL

A general aversion to genetically modified food has made the naturally caffeine-free Ethiopian strains that much more appealing. But breeding the trait into a commercially viable cultivar has taken longer than Mazzafera or Silvarolla anticipated. During the flowering season, Silvarolla spends all day in the field, clipping the pollen-producing stamens off the flowers and bagging the pistils so that she can hand-pollinate them later. She produces 800 new plants every year.

The beans that are produced taste good, according to tasting panels put together by the IAC, but the plants tend to be bushy and don't flower uniformly. The team is now working on breeding plants that have the low-caffeine trait but not the low-productivity one. "At first, we thought it would be a piece of cake," says Miriam Maluf, a geneticist at the IAC working on the project. But the researchers may be facing the biggest challenge of all in removing caffeine from living plants: it is there for a reason. Caffeine is a natural insecticide, which explains why wild coffee plants that lack caffeine tend to contain other bitter compounds — to deter pests.

And the researchers have to contend with these issues largely without Mazzafera's help. Soon after the 2004 discovery and subsequent publicity, the IAC took greater control of the programme, and

Mazzafera, based at the university, has only a limited role. "It's disappointing," he says.

Nevertheless, despite so many years without success, the quest for caffeine-free coffee shows little sign of fading. Plant geneticist Benoit Bertrand of the French Centre for Agricultural Research and Development in Montpellier has been searching the centre's collection for caffeine-free plants. And Chifumi Nagai of the Hawaiian Agriculture Research Center in Waipahu has been working in Madagascar with the Japanese coffee manufacturer UCC Ueshima to develop a three-species hybrid that tastes good, has a moderate yield and contains just 0.37% caffeine³. Its success is uncertain; Madagascar faces significant logistical challenges even growing and harvesting normal *C. arabica*.

Even Mazzafera, now 51, hasn't given up. On a cloudy November day, he walks out past two mesh shade houses behind the biology department in Campinas and reveals several hundred hip-high coffee plants. Some have rows of green coffee berries clustered on their branches. Many, he says, are nearly caffeine-free.

Miffed at his inability to continue his work with the Ethiopian varieties, he came up with a new plan in 2006. He took the seeds of a productive *C. arabica* variety, soaked them in chemicals that cause mutations, and then screened the caffeine levels of 28,000 seedlings. "It was a shot in the dark," he says. He ended up with 7 plants that have only 2% of normal caffeine levels⁴. He has already trademarked their name: Decaffito.

Challenges remain — the strain is susceptible to cross pollination, which can reinstate caffeine production in the beans — but he is determined to breed a commercially viable strain. He has even been talking to one company about investing in his research. But knowing the hurdles ahead, he is willing to settle for less. "If I had a farm," he says, "I would grow this coffee for myself." ■

The biggest challenge of all may be that caffeine is there for a reason.

Brendan Borrell is a freelance journalist in New York City.

1. Silvarolla, M. B., Mazzafera, P. & Fazuoli, L. C. *Nature* **429**, 826 (2004).
2. Ogita, S., Uefuji, H., Yamaguchi, Y., Koizumi, N. & Sano, H. *Nature* **423**, 823 (2003).
3. Nagai, C., Rakotomalala, J. J. & Katahira, R. *Euphytica* **164**, 133–142 (2008).
4. Mazzafera, P., Baumann, T. W., Shimizu, M. M. & Silvarolla, M. B. *Trop. Plant Biol.* **2**, 63–76 (2009).

COMMENT

MENTAL HEALTH The many advances that will hasten better drugs **p.267**

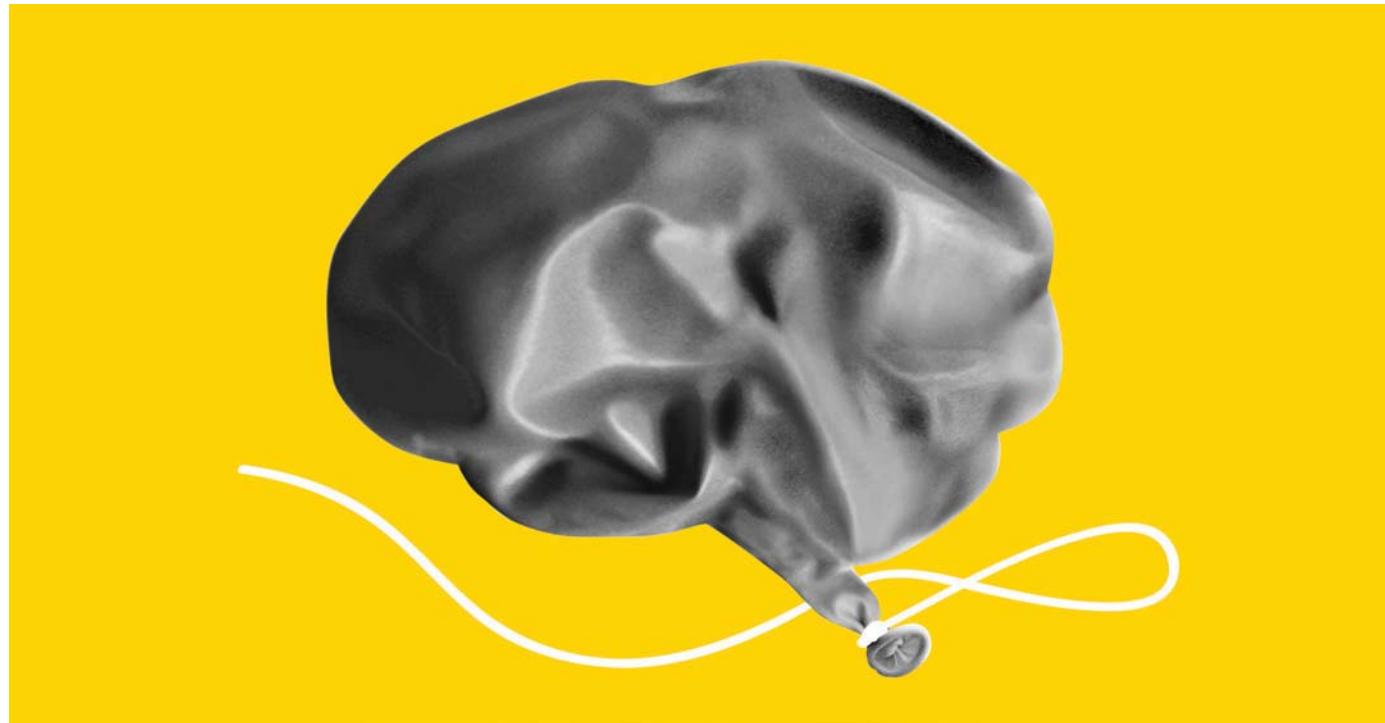


NEUROSCIENCE Exploring the origins of consciousness **p.271**

ANTARCTICA A portrait of the people who work across this frozen continent **p.272**

MATHEMATICS Eric Demaine on paper folding and geometry **p.274**

ILLUSTRATION BY O.O.P.S. FOR NATURE



Plug the real brain drain

Martin Schwab and **Anita Buchli** suggest ways to jump-start the stalled development of therapies for neurological diseases.

Recently, at the annual retreat of the Zurich Neuroscience Center, we ran into a former colleague who had often helped us to prepare for courses we were teaching. But he was not there to teach — he was participating in a demonstration as a patient. A stroke had left him paralysed on one side, wheelchair-bound and unable to speak. He had been looking forward to interacting with the students, but when he could not communicate with them, he broke into tears.

After a difficult rehabilitation, he was able to resume some of his work, but he still cannot speak. His arm and leg will probably remain paralysed for the rest of his life.

Our colleague was one of the 8.2 million Europeans who experience stroke every

year¹. The brain is a source of many devastating disorders — such as Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis — and injuries to the spinal cord or brain can lead to lifelong impairments. At present, disabling spinal-cord injuries affect roughly 350,000 people in Europe and 250,000 in the United States. Traumatic brain injuries are about ten times more common.

Treatments that could restore lost functions to people with such injuries would radically change their lives and decrease the burden

to their families and social environment. The economic interest to drug companies and health insurers seems obvious. Yet

drug companies have withdrawn from neuroscience, more so than from any other disease area. Last year, Novartis closed its preclinical neuroscience research facility in Basel, Switzerland. Pfizer, GlaxoSmithKline and AstraZeneca had already made similar moves. Merck and Sanofi are also cutting research on brain diseases.

Until recently, industry funded nearly half the budget for research and drug development for brain disorders². Its retreat has left an abyssal hole.

The reason for companies' reluctance to pursue drugs for neurological disorders is fairly straightforward: their investments haven't paid off. In the past 10–15 years, dozens of clinical trials for stroke ▶

ONATURE.COM

Read more about the crisis in brain-drug development:
go.nature.com/5goxwg

► neuroprotection — involving thousands of patients — have failed.

To get drug development going again, we must tackle the problems that have stalled it in the past by building a culture of interdisciplinary exchange to generate promising compounds and setting aside public funds to conduct small, well-designed clinical studies of those compounds. We realize that in such a tight funding situation, every field is asking for more. But given the extraordinary burdens neurological diseases cause, they must become more of a priority.

A NEW HOPE

Drug companies have pulled out of neuroscience just as our understanding of brain plasticity has exploded. The antiquated view of the central nervous system as a hard-wired supercomputer has been overturned; the brain and spinal cord now appear as dynamic and adaptable biological systems.

Large injuries to the brain and spinal cord are not repaired spontaneously, causing lifelong impairment. But scientists have recently developed experimental interventions that enhance nerve-fibre growth and regeneration in animals with massive brain injury. In experiments with rats, mice and monkeys, researchers (in our laboratory and others) have induced regrowth of injured nerve fibres in the brain and spinal cord by suppressing growth inhibitors — enough for the treated animals to regain lost functions^{3,4}.

We and our colleagues at Novartis are now conducting a clinical trial in which people with spinal-cord injuries receive an antibody that counteracts the neural growth inhibitor Nogo-A (also known as reticulon-4A). Other clinical trials to enhance repair of the spinal cord and brain are or will soon be under way. But progress is slow — the biotechnology company Geron, for example, recently abandoned a promising phase I trial of stem cells in spinal-cord injury to concentrate instead on cancer⁵.

Why have so many trials failed, and what should be done better? A drug may be effective and still fail in a trial. One reason is that companies often look for the most broadly applicable drug — for example, 'for all stroke patients' — but disease conditions often differ among patients, resulting in huge variations in treatment responses. Another problem with past trials was that the often crude clinical endpoints missed small but meaningful treatment effects, such as improvements in hand, leg or bladder function. With novel approaches, we can do better.

To reinvigorate the field and avoid repeating past problems, more exchange should

COSTS AND RESEARCH FUNDING IN EUROPE

Brain disorders are costlier and more prevalent than cancer but got similar research funding in 2005.

	Brain disorders	Cancer
Total costs for 2010 (ref. 1)	€798 billion (US\$1 trillion)	€150 billion–250 billion
Direct costs [†]	60%	41%
Indirect costs [‡]	40%	59%
Proportion of all health-care expenditures	24%	6.3%
Percentage of 2004 disease burden ⁷	35%	16.7%
Total 2005 research funding ²	€4.1 billion	€3.9 billion
Industry contribution	€3.3 billion	€2.5 billion

[†]Expenses such as medication or doctor's visits. [‡]Productivity lost to time off work or early retirement.

be fostered between basic and clinical scientists. When spinal-cord researchers began organizing retreats and workshops to bring together basic researchers and clinicians, they saw first-hand how little each side knows about how the other works. The mutual lack of knowledge was huge; each side had completely different language to describe the same scenario.

A WISER APPROACH

If researchers collaborate from the outset, they are more likely to produce a drug that works⁶. For instance, they could establish a set of criteria to evaluate a particular therapy in both animals and humans, so that what seems to work in one is more likely to seem to work in the other. Newer diagnostic tools will enable scientists to identify which subgroups might benefit most from a specific therapy. In addition, clinicians are now standardizing observations of functional improvement so that they can spot subtle changes that would have gone unnoticed in the past.

Neuroscience faculties and medical centres must work together to establish research consortia and networks that unite basic and clinical scientists. On a smaller scale, retreats with select groups of experts from both sides are inexpensive and can jump-start a field. Already, studies of spinal-cord injury are more focused now that the two sides are communicating — some basic researchers have begun using clinical criteria for functional improvement.

We can't just throw money and resources at the problem — we must use them wisely. Instead of investing billions in one drug, let's spread funding among smaller, proof-of-concept trials for compounds with good preclinical evidence. By focusing on well-selected populations (with tens of patients, not hundreds) and concentrating on a few centres, such trials would cost a few million euros rather than the €50 million (US\$67 million) or more needed for one large trial. If smaller trials can bring a promising compound to an advanced stage, industry may then be willing to take it to market.

And, because the pharmaceutical industry isn't ready to invest in early-stage research in

neurological diseases, we must turn to other sources. Insurance companies spend up to €2 million for each patient with a spinal-cord injury — a drug that could lower a patient's disability would save insurers huge amounts. In 2009, the top five US health insurers earned more than US\$12 billion; investment of even a small percentage of these profits in research could result in a true win-win situation.

It is in countries' best interests to dedicate more public money to small trials of therapies for brain diseases. People with such disorders may spend decades of their lives disabled, which can have enormous effects on their lives and on those of the people around them.

In 2011, a report commissioned by the European Brain Council found that, in terms of health-care costs and lost productivity, brain disorders are a greater socio-economic burden than cancer, cardiovascular diseases and diabetes combined¹. Yet in 2005, research funding for cancer and neurological diseases was roughly equal (see 'Costs and research funding in Europe'). More than half of that total comprised private funding; now that drug companies have shifted focus, cancer funding is likely to eclipse that of neuroscience. Funding agencies must revise their budgets to reflect the immediate and future needs of our society. ■ **SEE COMMENT P.269**

Martin E. Schwab and Anita D. Buchli
are at the Brain Research Institute of the University of Zurich and the Department of Health Sciences and Technology at the Federal Institute of Technology, CH-8057 Zurich, Switzerland.
e-mail: schwab@hifn.uzh.ch

1. Gustavsson, et al. *Eur. Neuropsychopharmacol.* **21**, 718–779 (2011).
2. Sobocinski, P., Lekander, I., Berwick, S., Olesen, J. & Jönsson, B. *Eur. J. Neurosci.* **24**, 2691–2693 (2006).
3. Zoerner, B. & Schwab, M. E. *Ann. NY Acad. Sci.* **1198**, E22–E34 (2010).
4. Liu, K., Tedeschi, A., Park, K. K. & He, Z. *Annu. Rev. Neurosci.* **34**, 131–152 (2011).
5. Baker, M. *Nature* **479**, 459 (2011).
6. Kwon, B. K., Hillyer, J. & Tetzlaff, W. *J. Neurotrauma* **27**, 21–33 (2010).
7. Andlin-Sobocinski, P., Jönsson, B., Wittchen, H. U. & Olesen, J. *Eur. J. Neurol.* **12**, 1–27 (2005).

A plan for mental illness

Thomas R. Insel and Barbara J. Sahakian describe what the next decade of mental-health drug development should look like.

ILLUSTRATION BY OOPS FOR NATURE

No chronic disease burdens the world more than mental illness¹. In any given year, nearly 40% of the population in 30 European countries is affected. And yet, the world is experiencing a crisis in drug development for mental illness; drug companies are withdrawing from the field or redirecting their investments². Last November, an eagerly awaited compound, called TC-5214, failed to significantly relieve major depression in a phase III trial³. The timing could not have been worse. Recognizing that the pipeline was starting to run dry, the Royal Society in London recently convened an international scientific seminar to find solutions to this pharmacological impasse. The meeting concluded that a fundamental change was needed in nearly every aspect of translational research in mental health.

For better treatments, better science is needed. Genetic studies are already identifying new molecular targets in other fields — for example, research in sickle-cell anaemia has uncovered a gene that represses fetal haemoglobin production, providing promising new drug targets⁴. The molecular mechanisms of mental disorders are very complex, but the latest genetic technologies should help to identify effective treatment targets, and offer alternatives to the monoamines that have dominated psychiatric medications for half a century.

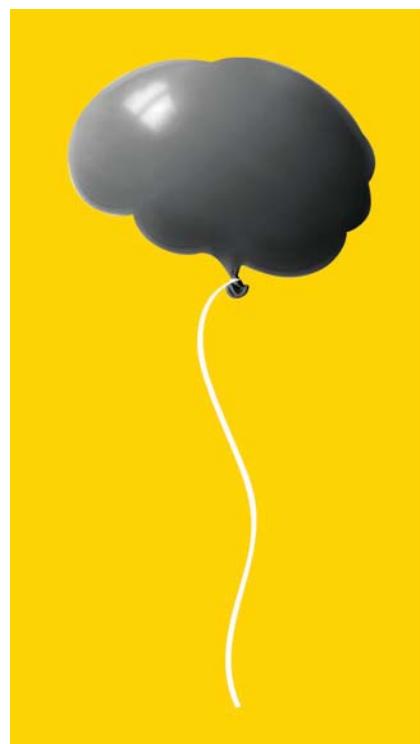
Borrowing from the success of prevention in other medical areas, ways to identify risk factors of mental illness need to be developed so that intervention precedes disability. Identifying these factors at an early stage may mean psychosocial treatments (such as, family support and cognitive training) are more effective than medication. Early detection could emerge from neuro-imaging, cognitive assessments, predictive biomarkers or rare genetic variants that signal risk or a prodromal stage.

Even therapeutic end points need to be refocused. Currently, treatment is not aimed at full recovery, and targets late stages of illness rather than using pre-emptive therapies in early stages when outcomes might be better⁵. Treatment should also address adherence, which remains a problem in 20–70% of patients with psychiatric illness⁶.

One method for improving the effectiveness of treatment is integrating medical and psychosocial approaches. In a study of phobias, which combined the drug D-cycloserine and cognitive-behavioural therapy, people

given medication during virtual-reality sessions designed to accustom them to fearful situations saw symptoms improve faster than those just given a placebo with virtual-reality sessions⁷. Other potential techniques and 'neurotechnologies' include video-game training to help people in the early stages of schizophrenia to retain function or to help children with autism to interact socially, for example by increasing eye contact⁸.

From a research perspective, progress requires changes not only in what we do but



also how we do it. As industry withdraws from the field of mental health, we have an opportunity to rethink how academia and industry can work together for the public good. One obvious way to do this is to allow academic scientists to investigate compounds that have been abandoned by industry. But we can not simply repeat the best efforts of a company-directed failed clinical trial without new strategies. We should consider new approaches to identifying potential drug targets and their validation — such as open-access drug development — to include industry and academia.

We must also reconsider our use of animal models; too many compounds that

seem to treat a disease in animals end up having little impact in humans. Although animal models may be unable to model all disease symptoms — such as those of schizophrenia — they can help to understand the neural mechanisms of specific symptoms, such as impulsivity⁹. Effective treatment of impulsivity could have a significant clinical effect on neuropsychiatric disorders such as attention deficit hyperactivity disorder, mania and substance abuse. We need to recognize the promise and limitations of animal studies, so trials can move quickly into work with human cells and volunteers to define or eliminate targets for drug development.

To achieve this vision, we need a new cadre of clinical neuroscientists⁸ to focus on integrating basic and clinical neuroscience, translational medicine, and new methodologies. Patients are becoming more involved in drug development through advocacy groups and allowing their information to be held in registries, providing research partnerships with an opportunity to identify biomarkers and specific phenotypes that pinpoint the patient subgroups most likely to benefit from a treatment^{9,10}.

The increasing public-health needs of those with mental illness, including chronic disability, high suicide rates and early death, demand that we do better in the next decade. ■

Thomas Insel is at the National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, USA. **Barbara J. Sahakian** is in the Department of Psychiatry and MRC/Wellcome Trust Behavioural and Clinical Neuroscience Institute, University of Cambridge, UK. The full author list can be found at the online version of this article (see go.nature.com/t1ihrn).
e-mail: bjs1001@cam.ac.uk

1. Collins, P. Y. *et al.* *Nature* **475**, 27–30 (2011).
2. Cressey, D. *Nature* <http://dx.doi.org/10.1038/news.2011.367> (2011).
3. Ledford, H. *Nature* **479**, 278 (2011).
4. Sankaran, V. G. *et al.* *Science* **322**, 1839–1842 (2008).
5. Sahakian, B. J., Malloch, G. & Kennard, C. *Lancet* **375**, 1854–1855 (2010).
6. Julius, R. J., Novitsky, M. A., Jr & Dubin, W. R. *J. Psychiatr. Pract.* **15**, 34–44 (2009).
7. Ressler, K. J. *et al.* *Arch. Gen. Psychiatry* **61**, 1136–1144 (2004).
8. Lehner, T. & Insel, T. *Acad. Psychiatry* **34**, 87–89 (2010).
9. Ledford, H. *Nature* <http://dx.doi.org/10.1038/news.2011.324> (2011).
10. Saha, K. & Hurlbut, J. B. *Nature* **478**, 312–313, 2011



The brain is full of feedback loops and neuronal links that combine to give rise to consciousness.

NEUROSCIENCE

The mind mapped

Robert Stickgold revels in a lively account of a quest to quantify consciousness.

The psychologist Stuart Sutherland wrote that it is impossible to define consciousness "except in terms that are unintelligible without a grasp of what consciousness means ... Nothing worth reading has been written about it." It is arguable whether Christof Koch's *Consciousness* provides such a definition, but the book is definitely worth reading.

Koch, chief scientific officer at the Allen Institute for Brain Science in Seattle, Washington, is perhaps best known for his work with the late Francis Crick, searching for the neurobiological 'correlates of consciousness'. Here, he succinctly lays out the story of that quest. Focusing on how the brain might produce the mind, Koch mixes descriptions of major experiments with self-reflection and

warnings of the inherent danger of the exercise.

From Koch's collaborations with Crick, whom he seems to idolize, to his struggles with religion and free will, this is an engaging mixture of personal anecdote, scientific fact and pure speculation. It is often charming: Chapter 2, for instance, is entitled, 'In which I write about the wellsprings

of my inner conflict between religion and reason, why I grew up wanting to be a scientist, why I wear a lapel pin of Professor Calculus, and how I acquired a second mentor late in life.'

For many, the richest parts of the book will be Koch's lucid descriptions of experiments such as his work with Itzhak Fried, a neurosurgeon who implanted electrodes into the hippocampi of people with epilepsy. In one patient, Fried found a single neuron that responded only to the name or pictures of Saddam Hussein; in another, he found one that responded only to pictures of the actress Jennifer Aniston. In a descriptive tour de force, Koch explains that although Fried dubbed these cells concept neurons, we can think of them as "the cellular substrate of the Platonic Ideal of Jennifer Aniston".

Koch discusses the two theories that form the cornerstones of his own views of consciousness. The first, which he proposed with Crick, describes the neural correlates of consciousness as involving "reverberatory feedback loops" (which link higher-order sensory regions to the prefrontal cortex) and corticothalamic loops (which link these regions and nuclei in the thalamus). These networks then marinate in a bath of neurochemicals released by fibres coursing up from a variety of brainstem regions. How these connections produce consciousness is explained by the second theory, the integrated information theory of consciousness.

Here, the story seems to stumble. Consciousness, Koch argues, is not an emergent property of the brain. Rather, it is inherent in all matter, as "the amount of integrated information generated by the system in [a given] state above and beyond [that] generated by its parts". If this seems unclear, it is. Koch states that calculating this quantity is extremely difficult, and demands "heuristics, shortcuts, and approximations". The higher the value, he says, the more conscious the system. But what does it mean to say that one system



PASIEKA/SPL

Consciousness: Confessions of a Romantic Reductionist
CHRISTOF KOCH
MIT Press: 2012.
184 pp. \$24.95,
£17.95

► has twice as much consciousness as another?

There can also be too much gee-whiz science for my taste. For instance, talking about the nervous system of the worm *Caenorhabditis elegans* with its 302 neurons, Koch describes the creature's "state of consciousness" as the current position of the system on a 302-dimensional graph, with 2^{302} possible states (for each combination of the 302 neurons either firing or not). The neural network is a material thing, Koch argues, but the shape of the graph describes the conscious experience of the worm at that moment, "its phenomenal experience".

By this argument, the 32,768-byte computer file holding this review has a consciousness defined by its position on a 32,768-dimensional graph with $256^{32,768}$ (or, put more quaintly, 1 followed by 87,000 zeros) possible states. Personally, I don't think any of them have much chance of being conscious.

There are other points on which *Consciousness* lacks clarity. Koch tackles most of the big questions: are mice and fruitflies conscious (yes); can computers be conscious (yes); are they already (yes); is there free will (not really); can we ever understand consciousness (yes); can the mind affect our behaviour (no); and even is there a god (yes, but absent since the Big Bang).

"What does it mean to say that one system has twice as much consciousness as another?"

Unfortunately, it is often hard to decipher which of these he thinks are proven facts, which personal beliefs and which momentary opinions.

In fairness, the equations are much more complicated than is described here. Still, in the end, the theory seems to me to be an example of a conflation of ignorances: saying that because we don't really understand A or B, they must be causally related. Whether they invoke quantum mechanics, microtubules or integrated information content, these explanations of consciousness are, for all their mathematical precision, no more than pure speculation. Perhaps one of them is actually right. But perhaps, like ancient Greek philosophers searching for the cause of thunder, we are still thousands of years from an answer.

I argued with Koch all the way through this book. And I loved every minute of it. ■

Robert Stickgold is an associate professor of psychiatry at Harvard Medical School in Boston, Massachusetts. e-mail: rstickgold@hms.harvard.edu



Humans have been probing the mysteries of the Antarctic for almost 200 years.

POLAR RESEARCH

Deep-frozen science

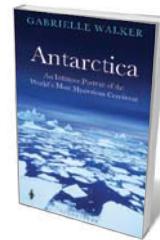
Francis Halzen is exhilarated by a trek through stories of research and exploration in Antarctica.

Adazzling array of narratives throngs *Antarctica*, collected from scientists in one of Earth's most extreme environments. Science writer and consultant Gabrielle Walker gathered these stories in the course of five trips criss-crossing the continent, mostly as a guest of the US National Science Foundation's Antarctic programme.

This is not just a highly accessible encyclopedia of Antarctic science. It interlaces researchers' stories with natural history, tales of the 'heroic age' of exploration and passages that viscerally describe the cold, isolation and beauty of the environment. Neatly organized geographically, the book covers the East Antarctic coast, with McMurdo Station (the largest community in Antarctica) and the penguins; the high plateau, with the Concordia and South Pole research stations; and the isolated West Antarctic coast.

Walker talks about the Dry Valleys near McMurdo, which run from the edge of the ice sheet that covers most of the continent down to the Ross Sea. Precious little precipitation has fallen here for millions of years: with an average temperature of -55°C , this is Mars on Earth. But it is teeming with life — cyanobacteria, found in ponds the world over. Here, they live inside the dry rock, surfacing for only a few weeks a year to find water from the little snow that fell over winter. Then they go back to sleep.

Walker learned that fascinating story on a diving expedition in Lake Hoare in the Dry Valleys with Peter



Antarctica: An Intimate Portrait of the World's Most Mysterious Continent
GABRIELLE WALKER
Bloomsbury: 2012.
416 pp. £20

Doran, a biologist at the University of Illinois at Chicago. But it isn't only researchers who feature here. The book is peppered with characters who "keep the scientists alive". On one trip we meet camp manager Rae Spain, who came to Antarctica as a carpenter and returned because she could not get the continent out of her head. "It haunts you," she told Walker.

Mars has literally come to Earth just beyond the Dry Valleys, Walker tells us. Researchers on Skidoos systematically scan the ice sheet for meteorites — a relatively easy task, given that everything stands out in this icy landscape. We accompany John Schutt, a mountaineer who has returned for the hunt every year since 1980. More meteorites have been found since the 1970s in Antarctica than over centuries in the rest of the world.

In 1982, for the first time, an Antarctic meteorite was identified as coming from the Moon. Two years later, researchers found a rock from Mars that turned out to contain structures that may be nanoscale fossils: the most intriguing indications yet that life may have existed on other planets in the Solar System, although debate is still raging.

The West Antarctic coast, as Walker shows, is out of reach of the permanent

scientific stations, and difficult to approach by vessel. This is where gigantic glaciers empty into the Pacific Ocean, and studies have revealed rapid changes in the ice flow. Helen Fricker, a glaciologist at the Scripps Institution of Oceanography in La Jolla, California, has traced these changes to subglacial lakes linked by canals, which form a dynamic hydrologic system on which the ice slides to the sea. Walker journeys across the inhospitable glaciers, where she stands on ice flows the size of Niagara Falls with George Denton of the University of Maine in Orono, a veteran in 'reading' glacial landscapes.

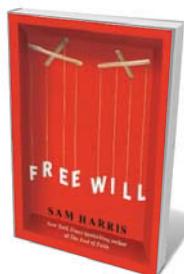
Some of the historical vignettes will be familiar, such as the race to the South Pole between Norwegian explorer Roald Amundsen and his trained team, and Robert Scott, British hero of scientists, who presumably found Amundsen's efficient approach ungentlemanly. Other stories are less well known, such as that of Australian geologist Douglas Mawson, who travelled to the south magnetic pole as part of Ernest Shackleton's Nimrod expedition and later led a disastrous research trip to Adelie Land, on which two men died.

Tales of those expeditions, and the scientific ones that followed, remind us that people have been travelling to Antarctica for almost two centuries, ever since seal hunter Captain John Davis first set foot there in 1821. The Antarctic Treaty, signed by 49 countries, has since 1961 guaranteed that the continent is reserved for science. High technology has now arrived: the collaborations behind the South Pole Telescope and the IceCube Neutrino Observatory have constructed the kinds of instrument that are more routinely built at laboratories such as Fermilab near Batavia, Illinois, or CERN near Geneva, Switzerland. IceCube collects a few hundred neutrinos per day, some with energies that exceed those at earthbound accelerators by more than two orders of magnitude.

Yet Antarctica is still the "world's most mysterious continent", as it remains the only one on which humans have never lived permanently. Walker captures that mystique through interviews with people who have made Antarctica part of their lives. Perhaps the most notable among them are the "telescope nannies", who return each year in early February to spend the long Antarctic winters at the South Pole, taking care of the scientific equipment and data acquisition after scientists have boarded the last planes back to their universities. Their arduous job brings a rare reward: a sense of the untrammelled isolation of this vast continent. ■

Francis Halzen is the principal investigator of the IceCube project, and Hilldale and Gregory Breit Professor in the department of physics at the University of Wisconsin-Madison.
e-mail: halzen@icecube.wisc.edu

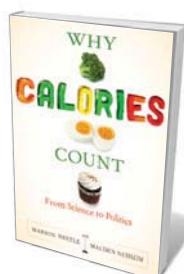
Books in brief



Free Will

Sam Harris FREE PRESS 96 pp. \$9.99 (2012)

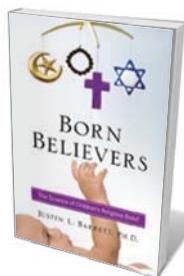
Neuroscientist Sam Harris, the author of the bestselling *The Moral Landscape* (2010), here skewers the concept of free will — that mainstay of law, policy and politics — in fewer than 100 pages. Using evidence drawn from psychology and neuroscience, Harris asserts that the course of human life is all down to luck, and that willpower is a "biological phenomenon". We are not in charge of our own minds, he says: thought and intention simply arise. This is a tract that is sure to boldly go straight to the heart of the determinism debate.



Why Calories Count: From Science to Politics

Marion Nestle and Malden Nesheim UNIVERSITY OF CALIFORNIA PRESS 304 pp. \$19.95 (2012)

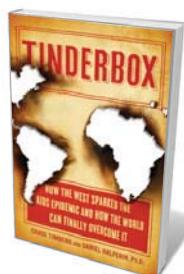
Obesity has gone global — as has misinformation about nutrition and food. Nutrition scientists Marion Nestle and Malden Nesheim unscramble the confusion with a serving of science. They reveal how calories — those potent but ill-understood measures of heat energy — are really counted, why we need them, how we use them, how many we actually need and why it all sometimes goes so wrong. From 'secret' calories to food politics, malnourishment and calorie restriction for health, this is a feast for the mind.



Born Believers: The Science of Children's Religious Belief

Justin L. Barrett FREE PRESS 320 pp. \$26.00 (2012)

Psychologist Justin Barrett says that belief is rooted not in indoctrination, but in a default setting of the infant brain. A range of findings in developmental psychology, he argues, support the theory: very young children, for instance, are aware that nature is not designed by humans; and people tend to look for unseen agents behind natural phenomena. Barrett says that the evidence points to a universal "natural religion" — but he also argues that parents must not impose belief systems on their children, and should leave the question of whether to die a believer up to their offspring.



Tinderbox: How the West Sparked the AIDS Epidemic and How the World Can Finally Overcome It

Daniel Halperin and Craig Timberg PENGUIN 432 pp. \$29.95 (2012)

The butchering of a chimpanzee in Cameroon about a century ago may have been when simian immunodeficiency virus crossed into humans and became HIV. But, argue epidemiologist Daniel Halperin and journalist Craig Timberg, European greed and technologies sparked the HIV pandemic. Trade routes, the growth of colonial cities, a decrease in male circumcision and a rise in prostitution effectively incubated and spread the disease — and Western programmes to curb it in Africa have proved mostly ineffectual, say the authors.



Digitized: The Science of Computers and How it Shapes Our World

Peter J. Bentley OXFORD UNIVERSITY PRESS 256 pp. £16.99 (2012)

The wiring of the world — from a glint in Alan Turing's eye to global domination — is neatly encapsulated by computer scientist Peter Bentley. An enthusiastic guide, he lays this compelling history bare, explaining the science as he goes: Moore's law of processing-power growth, the circuit designs of Claude Shannon, John von Neumann's early electronic digital computer, neural networks and the staggering array of applications for computing in everything from pizza delivery to Alzheimer's-disease research.

Curved-crease origami sculptures can self-fold into intricate patterns.



Q&A Erik Demaine

The origami geometer

Computer scientist Erik Demaine uses origami to advance computational geometry and create art. His paper sculptures, made with his father, artist Martin Demaine, are now on show at the Japanese American National Museum in Los Angeles, California; from August, the exhibition will tour the United States. He explains the challenges of folding together mathematics and art.

NICK HIGGINS

How did you come to do most of your work with your father?

When I was five, I helped him to design wire puzzles for toy shops across Canada. He educated me himself as we travelled around the United States. When I went to university at the age of 12, he attended all my classes, mostly computer science and mathematics. Eventually, we started working together on mathematics and art. He is now an artist in residence at the Massachusetts Institute of Technology (MIT) in Cambridge, where I am a computer scientist.



Folding Paper: The Infinite Possibilities of Origami

Japanese American National Museum, Los Angeles, California.
March–August 2012.

making one straight cut, then unfolding it? Houdini made a five-pointed star in this way in the 1920s. It took us years to prove that you can make any straight-edged shape; we have designed swans, butterflies and the MIT logo. Another result was an algorithm for folding any three-dimensional (3D) shape out of one sheet of paper.

What is 'self-folding' origami?

It is a form in which the force of the creases pulls the material naturally into shape, without any manipulation. We're trying to find an algorithm to make any 3D shape with this method. One application would be a space station that assembled itself in space. We have devised a grid-based crease pattern that would allow a microscopic sheet to self-fold into any shape, in theory, by making cubes that stack together like 3D pixels. We haven't yet achieved self-folding at the nanoscale.

What kind of origami are you doing now?

We have been working with curved creases, the properties of which are poorly understood. We were trying out some curved patterns, and saw that

NATURE.COM
For more on maths and origami:
go.nature.com/hitdq

they looked beautiful. We're still pursuing unsolved mathematical problems too. David Huffman, who invented the compression algorithms used for mobile phones, left many beautiful origami sculptures when he died in 1999. He was getting to the point at which he could conceive a 3D form, then reverse engineer the steps to fold it with curved creases. We hope to have him as a posthumous co-author on some papers.

How do you actually fold the paper?

For prototype models we use a robotically controlled laser to score the paper, but we prefer a simple compass-like device. The final folding we always do by hand. At the first show we contributed to, we made three sculptures by folding two circular pieces of paper together using alternating pleats in concentric circles. Curved-crease sculptures go back to a late-1920s design by the Bauhaus school in Germany. The surprising thing is how they self-fold into ornate forms from these simple creases. For the *Folding Paper* exhibition, spanning the range of origami art, we gave ourselves a little more power by manually joining a few key points together.

How did you come to study balloon animals?

That was a collaboration with my father and Vi Hart, a composer whose own father is a mathematical sculptor. We noticed that balloon animals could be seen as outlining the edges of a flat-sided 3D solid, and we found an algorithm that tells you how many balloons you need to build a given solid. One application, conceived after the fact, would be to build a pavilion from a single tube. But our motivation was just to play with balloons. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Ads against chimp research criticized

We at the Federation of American Societies for Experimental Biology object to your publication of advertisements from the Humane Society of the United States that seem to misrepresent the importance of chimpanzees in biomedical research and testing (see, for example, facing page 407 in 22/29 December 2011 issue of *Nature*).

These advertisements claim that chimpanzees are a poor model for research into human disease because of their physiological and immunological differences. This contention is at odds with the findings of a US Institute of Medicine committee appointed to investigate whether chimpanzees are necessary for biomedical and behavioural research (see go.nature.com/ruthsl).

The committee, composed of scientific experts and led by a bioethicist, concluded that chimpanzees have been valuable models in the past and that, although "most current use of chimpanzees for biomedical research is unnecessary", they are still needed to conclude research on monoclonal antibodies and possibly in the development of a prophylactic hepatitis C vaccine.

The committee emphasized that an outright ban on biomedical chimpanzee research would not be appropriate, in part because of new, emerging or re-emerging diseases that cannot be studied in non-chimpanzee models.

Joseph C. LaManna FASEB, Bethesda, Maryland, USA.
Joseph-LaManna@faseb.org

Turing: Colossus computer revisited

George Dyson's history of the Colossus computer (*Nature* **482**, 459–460; 2012) is somewhat misleading. The development of Colossus owed little to the Bletchley Park 'bombe' devices.

After Bill Tutte developed a statistical strategy to tackle the German Fish cypher, Max Newman proposed a machine to implement it, using two paper tapes. With help from the Telecommunications Research Establishment (later the Royal Radar Establishment) and from Frank Morrell's group at the Post Office Research Station at Dollis Hill, the Robinson machines were built. These were slow but validated the technique. Thomas Flowers' group recommended the use of many more vacuum tubes to improve performance, but they were wrongly considered too unreliable by Bletchley Park.

So Flowers built the first Colossus at Dollis Hill instead of Bletchley Park although, in time, Newman spoke to him about his requirements.

One crucial technique developed by Turing and used by Tutte was that of 'delta-ing', or using the differences between characters rather than the characters themselves. If Dyson had pointed out that Tutte used this method to detect the statistical non-randomness of plain text, it would also have helped in understanding how different Fish was from Enigma.

Henry Shipley Hexham, UK.
h.h.shipley.61@cantab.net

Turing: Brain model still incomplete

In contemplating whether the brain is a good model for machine intelligence (*Nature* **482**, 462–463; 2012), I believe that Alan Turing's principle that the brain performs computations will continue to hold true. But it seems clear from the state of machine intelligence today that we're missing some basic insight into the language of brain computation.

Advances in genetics, electronics and optics are now enabling us to look into simple brains while they're living and behaving, and to observe

the activity of every neuron in real time. This influx of information should eventually help us to make intelligence comprehensible and replicable.

A good approach to interpreting these data is the re-creation of neural processing in simulation, from sensation all the way to behaviour. But once we understand what's going on, it should be possible to create machine intelligence that doesn't rely on reproducing low-level chemical dynamics.

Over the coming decades we are likely to take high-level cues from biology on how to organize our silicon. Mind is merely a function of the brain, however, so we should be able to capture that function in any Turing-complete system once we know how the brain is organized to perform it.

David Dalrymple Cambridge, Massachusetts, USA.
davidad@alum.mit.edu

Turing: Beyond the original concept

Barry Cooper asks whether information can be increased through computation, pointing out that Turing computation does not create anything not already in the initial data (*Nature* **482**, 465; 2012). If we are limited to Turing machines, then I believe the answer to his question is no. But if we enhance them then information can accumulate.

Turing machines are designed to model functions, not ongoing computations involving additional input over time. But if we enhance Turing machines by giving them a persistent memory and allowing them to alter their input by interacting with their environment, then information can increase.

Interactive machines engage in input and output during computation, which is closer to how computers are used in practice than in the Turing machine framework: examples include operating systems,

interactive agents in artificial intelligence, and solutions to some problems in control theory.

This observation is largely unappreciated in theoretical computer science, but a handful of researchers are exploring this part of 'super-Turing' space.

Christopher Kanan University of California San Diego, La Jolla, California, USA.
ckanan@cs.ucsd.edu

Be sparing with international laws

We believe that Devi Sridhar's justification for proposing a Framework Convention on Alcohol Control is problematic (*Nature* **482**, 302; 2012). Standard-setting international laws are largely dictated by powerful states, based on expectations that they themselves already meet, obliging poorer states to implement these 'enlightened' global policies ahead of local priorities. Litigation by foreign non-governmental organizations (NGOs) can also get in the way of national policy-making, particularly as most NGOs are led from the West.

Sridhar's proposal joins growing calls for new international health laws. But first we need more evidence that international laws achieve results commensurate with the cost of drafting, ratifying and implementing them.

In our view, some clear criteria need to be fulfilled before the WHO invokes its law-making authority. These could be set by a commission on global health law. Ill-justified international health laws that dictate poor countries' policies and priorities from afar could prevent serious consideration of initiatives better suited to legal instruments.

Steven J. Hoffman McMaster University, Hamilton, Canada.
hoffmans@mcmaster.ca

John-Arne Røttingen Harvard Kennedy School, Cambridge, Massachusetts, USA.

Catalysis in tight spaces

An acid has been found to catalyse the formation of a common chemical group, the spiroacetals, and to control which mirror-image isomer of the group is made. The key to success is the acid's bulky molecular structure. **SEE LETTER P.315**

NOAH Z. BURNS & ERIC N. JACOBSEN

Opposites attract. When applied to the charges in organic molecules, this idea serves as a foundation for understanding how such molecules react. A large portion of organic chemistry can be explained by the concept that electron-rich molecules (or sites on molecules), which have negative character, seek to interact with, and potentially donate electrons to, electron-poor species that have positive character. On page 315 of this issue, Čorić and List¹ describe a fresh use of this concept, with a report of a negatively charged, bulky catalyst that controls the reaction of a positively charged intermediate in the synthesis of potentially useful compounds known as spiroacetals.

Some of the most reactive, and therefore the most synthetically useful, species for organic transformations are those bearing a full negative or positive charge. Of these, oxocarbenium ions — oxygen-containing cations that form as reaction intermediates (Fig. 1a) — are particularly useful for making a variety of chemical linkages and groups. For example, they are used to make acetal groups, in which a carbon atom is connected to two oxygen atoms

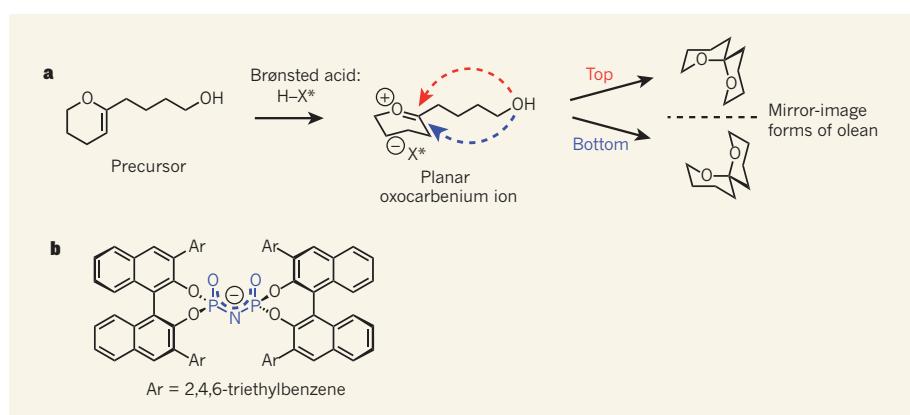


Figure 2 | A spiroacetal-forming reaction. **a**, Čorić and List¹ report a method for preparing chiral spiroacetals from a planar oxocarbenium ion. The ion is generated by treating its precursor with a Brønsted-acid catalyst, H-X. If the acid is not chiral, the alcohol group (OH) will attack the planar oxocarbenium ion from either face with equal likelihood, leading to a racemic (50:50) mixture of mirror-image isomers of a spiroacetal product (in this case, olean). **b**, With the chiral Bronsted acid developed by Čorić and List, the chiral anion X⁻ generated after protonation (the asterisk denotes that the anion is chiral) can selectively block one face of the oxocarbenium ion, generating one mirror-image isomer selectively. The mirror-image isomer of the catalyst that is chosen determines which isomer of the product is produced. The anion formed from the acid catalyst is shown; the broken blue line indicates that the negative charge is delocalized across the atoms shown in blue.

(Fig. 1b). Acetals are key linkages in numerous biomolecules, notably serving to bond sugar molecules together in long polymeric chains such as those found in amylose and amylopectin (the two components of starch), as well as in cellulose.

The precise stereochemistry (the three-dimensional spatial arrangement) of acetals has profound implications. For example, both amylose and cellulose are made from the same glucose monomer, but have markedly different physical, chemical and biological properties. All that differentiates these two compounds is the stereochemistry at the acetal linker. Therefore, finding ways to control acetal stereochemistry — precisely what Čorić and List report¹ in their reaction for constructing spiroacetals, a subclass of acetals formed when the acetal carbon joins two rings (Fig. 1c) — is a highly relevant challenge. The reaction that the authors report controls the chirality, or 'handedness', of the products that form.

Spiroacetals are found in a wide variety of naturally occurring, biologically active small molecules. For example, integramycin — a potent inhibitor of the integrase enzyme² that

allows HIV-1 to integrate its genetic material into the DNA of a host cell — contains a spiroacetal. Olean, the pheromone of the olive fruitfly, is one of the simplest spiroacetals (Fig. 1d), and provides another illustration of the importance of stereochemistry to a molecule's properties: the two mirror-image forms of olean selectively attract flies of a different sex³.

Čorić and List¹ constructed the spiroacetals from a planar oxocarbenium ion that contained a pendant alcohol group (Fig. 2a). The stereochemical challenge was to control the addition of the alcohol group to just one face of the planar ion. One solution would be to use stereochemistry already present elsewhere in the molecule to direct the selectivity of the addition. But such internal control is, by definition, substrate-dependent and rarely allows both possible stereochemical outcomes of the reaction to be made selectively. A more recent and potentially more general approach is to control the stereochemistry using an external, negatively charged catalyst that interacts directly with the oxocarbenium ion through Coulombic attraction^{4,5}. Ideally,

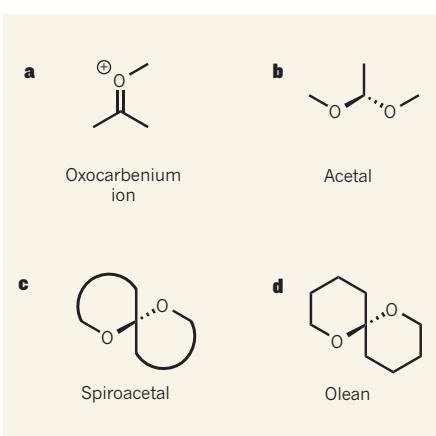


Figure 1 | Acetals and an acetal precursor.

a, Oxocarbenium ions are reactive intermediates that can be used to make various chemical groups, including **b**, acetals and **c**, spiroacetals. The rings depicted in the spiroacetal structure represent rings of any size. **d**, One of the simplest spiroacetals is olean, the pheromone of the olive fruitfly.

the stereochemistry of the catalyst alone can determine which of the two isomers of acetal is generated.

The problem with this seemingly straightforward strategy is that a very specific spatial relationship between the reacting site and the chiral catalyst is required to obtain high stereoselectivity. Electrostatic attraction alone is not sufficient to impose the required level of order, because it is direction-independent. This problem could be overcome by incorporating other features into the catalyst, to allow it to orient the charged substrate through attractive, non-covalent interactions — a principle that is well established in enzymatic catalysis, and which has also been established in small-molecule catalysis⁶. An alternative approach is to create a rigid, highly restricted environment near the charged area on the catalyst that might allow the cationic substrate to interact with the catalyst only in a highly specific manner.

Čorić and List¹ adopted the second of these strategies in their work. They used an exceptionally bulky Brønsted acid — a compound that can donate protons (H^+) to other molecules — as a catalyst⁷ to construct a variety of small spiroacetals in a highly stereoselective manner. In the authors' system, the catalyst donates a proton to the substrate, which becomes an oxocarbenium ion. The deprotonated Brønsted acid (Fig. 2b) then acts as a counteranion for the oxocarbenium ion. A complementary approach^{8,9} for generating oxocarbenium ions is to use a catalyst that removes an anion from a suitably reactive substrate; by forming a negatively charged complex with the anion, the catalyst controls subsequent additions to the oxocarbenium ion.

A notable aspect of Čorić and List's work is that simple substrates can be used that lack the large or polar groups often required in Brønsted-acid-catalysed reactions to ensure effective catalyst–substrate interactions. The authors propose¹ that this is because they used a particularly large Brønsted-acid catalyst, so the bulky anion that forms from the catalyst after the proton-transfer step can interact in a specific enough manner with the oxocarbenium ion to preferentially block one of the cation's faces. This prevents the alcohol from adding to that face and thereby directs addition selectively to the other face. Addition to either face can be selected for in a reaction simply by choosing the appropriate mirror-image form of the catalyst.

At present, Čorić and List's method seems to be limited to reactions that generate five- or six-membered spiroacetal rings. One remaining question is whether the success of the catalyst with simple substrates may be extended to more complex molecules. For example, the authors report¹ that their Brønsted acid is unable to control and override the inherent selectivity of spiroacetal formation in a large steroid molecule that contains many other

chemical groups and rings of atoms. Many biologically active spiroacetals are similarly complex, and so a method for synthesizing them would be highly desirable.

A mechanistic investigation of Čorić and List's reactions will be needed to establish whether the proposed confined-space effect is in fact operative, and to discover precisely how the cation–anion pair formed during the reactions yields spiroacetals stereoselectively. But such effort seems worthwhile, given the potential of bulky Brønsted acids as catalysts for other valuable organic transformations. ■

Noah Z. Burns and Eric N. Jacobsen are in the Department of Chemistry and Chemical

Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: jacobsen@chemistry.harvard.edu

1. Čorić, I. & List, B. *Nature* **483**, 315–319 (2012).
2. Singh, S. B. *et al.* *Org. Lett.* **4**, 1123–1126 (2002).
3. Haniotakis, G., Francke, W., Mori, K., Redlich, H. & Schurig, V. *J. Chem. Ecol.* **12**, 1559–1568 (1986).
4. Mayer, S. & List, B. *Angew. Chem. Int. Edn* **45**, 4193–4195 (2006).
5. Hamilton, G. L., Kang, E. J., Mba, M. & Toste, F. D. *Science* **317**, 496–499 (2007).
6. Knowles, R. R. & Jacobsen, E. N. *Proc. Natl. Acad. Sci. USA* **107**, 20678–20685 (2010).
7. Akiyama, T. *Chem. Rev.* **107**, 5744–5758 (2007).
8. Zhang, Z. & Schreiner, P. R. *Chem. Soc. Rev.* **38**, 1187–1198 (2009).
9. Reisman, S. E., Doyle, A. G. & Jacobsen, E. N. *J. Am. Chem. Soc.* **130**, 7198–7199 (2008).

ELECTRONICS

Inside story of ferroelectric memories

The drive to improve digital memory through ever-shrinking electronic circuitry will ultimately face a bottleneck. Researchers propose exploiting the room 'inside' memory elements as a solution.

VINCENT GARCIA & MANUEL BIBES

The amount of stored information in the world is reaching an astronomical 5×10^{21} bits — roughly one million times the number of cereal grains produced on Earth in a year². And the number of bits doubles every three years¹. This has motivated efforts to produce improved digital memory solutions that have better reliability, lower power consumption, lower cost and,

crucially, higher storage density. Writing in *Advanced Materials*, Lee *et al.*³ propose a method to increase the storage density of an existing class of commercial memory. The technique involves controlling the physical process through which information is written in memory cells — the building blocks of data-storage devices — instead of the usual approach of reducing the cells' size.

The data-storage market is dominated by hard-disk drives (HDDs) and Flash memories.

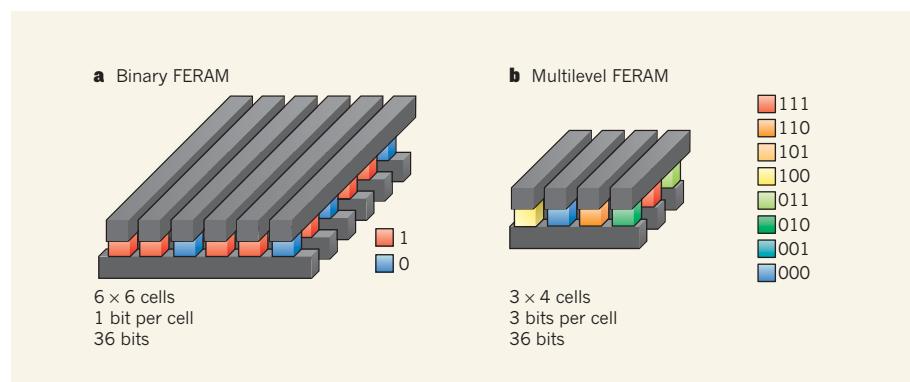


Figure 1 | Binary and multilevel data storage. **a**, In a classical binary ferroelectric random access memory (FERAM), a bit of information, with a logic value of '1' or '0', is encoded in the up (red block) or down (blue block) polarization direction of the device's constituent cells. In this 36-bit example, the bits are stored in 6×6 cells, each of which stores one bit. Electrical wires (grey) are used to apply voltages across the cells to write and read the cells' polarization. **b**, Lee *et al.*³ describe a multilevel FERAM that can store three bits in a single cell. In this way, three-bit logic states (000, 001, 010, 011, 100, 101, 110 and 111) can be stored in each of 3×4 cells of a smaller 36-bit memory array. Each state corresponds to a specific configuration of domains of opposite polarization in the cells.

DEVELOPMENTAL BIOLOGY

A brainy background

The vertebrate brain is so complex that tracing its origins among invertebrates is problematic. It must have evolutionary roots somewhere, but where? On page 289 of this issue, Pani *et al.* point the way with their finding that acorn worms have gene-expression signatures very similar to those that direct vertebrate brain development (A. M. Pani *et al.* *Nature* **483**, 289–294; 2012).

The closest relatives of vertebrates are the urochordates (tunicates, or sea squirts) and cephalochordates (the superficially fish-like lancelet). More distant on the evolutionary tree lie the hemichordates, which include acorn worms such as *Saccoglossus kowalevskii* (pictured), and the echinoderms (including starfish and sea urchins). Together, these groups comprise the deuterostomes, among which the elaborate structural and genomic innovations of vertebrates stand out rather like a cuckoo chick in a nest of willow warblers.

One such feature is the signalling

centres — physical regions of a developing embryo in which a specific set of secreted proteins drives the formation of a particular body section. These centres are present in the vertebrate neuroectoderm, an embryonic tissue that gives rise to the brain and nervous system. They are modified or completely absent in tunicates and lancelets, suggesting that many aspects of brain development are unique to vertebrates and arose *de novo* in the vertebrate lineage.

Pani and colleagues' data contradict this theory. They show that a gene-expression program similar to that in three vertebrate neuroectodermal signalling centres is also present in *S. kowalevskii*. Although acorn worms do not have anything like a brain, Pani *et al.* show that these genetic programs are used to pattern (direct the development of) the ectoderm of the embryonic animal. The genes expressed are equivalent to those found in vertebrates, and they are deployed at similar times and places in vertebrate and acorn-worm embryos.



A. M. PANI

The researchers propose that these signalling centres were part of an ancient gene-regulation 'scaffold' that was present in the common ancestor of all deuterostomes, and that was retained in hemichordates and vertebrates but lost in the evolutionary branches that formed lancelets and urochordates. Such an idea implies that the common ancestor was unexpectedly complicated. This does not mean that it had a complex brain, but that the genetic programs that were eventually modified for use in patterning the vertebrate brain already existed in an ancestral creature that lived perhaps more than 600 million years ago. **Henry Gee**

HDDs provide the highest-capacity solutions, with an unbeatable cost per bit. But they are slow and prone to failure when shocked, causing the loss of documents, data or holiday pictures. Solid-state memories such as Flash, which are ubiquitous in memory sticks and memory cards for cameras, are faster and more robust than HDDs. However, they cannot match the cost per bit and maximum storage capacity of HDDs. But they are well suited to mobile electronic devices, and the consensus is that most future data-storage solutions will be based on solid-state memories.

The question is how to increase storage density and reduce the cost of solid-state memories while maintaining their good performance (fast data-access and write times, low power consumption and high endurance). Following the advice of physicist Richard Feynman, researchers have for decades exploited the "room at the bottom" — that is, they have tried to reduce the size of memory cells. Impressive progress in nanometre-scale lithography has allowed a marked reduction in the size of a memory cell. So today, the smallest commercial Flash memory cell⁴ has a lateral size of about 80 nanometres.

To further reduce the size of memory cells, problems related to charge leakage, which causes undesired data erasure, and general reliability issues must be solved. This is proving more challenging than ever as cells reduce in size. One solution is to increase storage density not by using fine-pitch lithography but by piling up two-dimensional arrays of

memory cells and creating three-dimensional structures, thereby exploiting the 'room at the top'. The practical implementation is, however, extremely difficult in terms of circuitry, which prevents the pile-up of more than a few layers.

In their study, Lee *et al.*³ propose another method of data storage that exploits 'the room inside' the memory cell itself to store more than just one bit of information. Rather than Flash, the authors address ferroelectric random access memories (FERAMs), which, along with phase-change and magnetic memories, are an alternative type of non-volatile random-access-memory technology⁵. In a FERAM cell, a bit of information (a '1' or '0' logic state) is usually stored in the direction (up or down) of the polarization of a ferroelectric material — one that can retain a permanent electric dipole in the absence of an electric field (Fig. 1a). The polarization of a ferroelectric is the sum of the electric dipoles present in each unit cell of the material. Its direction can be switched by applying a voltage, just as the magnetization in a ferromagnet (a material that maintains a magnetic dipole in the absence of a magnetic field) can be reversed with a magnetic field.

Rather than just relying on the direction of the polarization to store information in a FERAM cell, Lee and colleagues make the best of the physical mechanism at the heart of a memory operation by taking advantage of the structure of domains of different polarization orientation in the ferroelectric material.

When a voltage is applied to reverse a ferroelectric's polarization, it usually does not reverse all of the material's constituent electric dipoles homogeneously. Instead, small regions (the domains), each with their own polarization, nucleate and then progressively expand. Depending on the voltage value, or the time during which it is applied, configurations — in which up and down domains coexist — can be stabilized. The net polarization is the sum of the up and down domains, and thus intermediate states of net polarization may be used to store information. These intermediate states define a multilevel memory element.

However, owing to the stochastic nature of ferroelectric polarization switching, controlling the up- and down-domain fractions by this voltage-based procedure creates poorly defined final intermediate polarization states. Lee *et al.*³ show that this poor definition can be greatly diminished by limiting the electrical current that appears in the ferroelectric when its polarization starts to be reversed by the applied voltage. Thus, complete polarization reversal is hampered, and the ferroelectric is left in a multidomain configuration. In this way, the authors were able to write eight well-defined logic states (000, 001, 010, 011, 100, 101, 110 and 111; Fig. 1b), converting a one-bit conventional FERAM memory element into a three-bit one. Although the main results were obtained on single-crystalline ferroelectric materials, they also show that the approach is applicable to polycrystalline films, which are typically used in commercial FERAMs.

The storage density of today's FERAMs is tenfold lower than that of Flash⁶, but exploiting ferroelectric domains could greatly improve it. Along with their excellent power consumption, endurance and write times, such improved storage density would put FERAMs back in the race towards a universal memory. In addition, the use of multidomain structures to encode intermediate polarization states could be applied to other devices based on ferroelectric layers, such as ferroelectric transistors⁷, tunnel junctions⁸ or switchable

photovoltaics⁹. Lee and colleagues' work also emphasizes how a deep understanding of the physical processes underlying information storage allows even the highest technological hurdles to be overcome. ■

Vincent Garcia and Manuel Bibes are at Unité Mixte de Physique CNRS/Thales, Campus de l'Ecole Polytechnique, Palaiseau 91767, France.
e-mails: vincent.garcia@thalesgroup.com; manuel.bibes@thalesgroup.com

1. Hilbert, M. & López, P. *Science* **332**, 60–65 (2011).
2. <http://faostat.fao.org>
3. Lee, D. *et al.* *Adv. Mater.* **24**, 402–406 (2012).
4. Goldman, M. & Pangal, K. *3rd IEEE Int. Memory Workshop* <http://dx.doi.org/10.1109/IMW.2011.5873197> (2011).
5. Fujisaki, Y. *Jpn. J. Appl. Phys.* **49**, 100001–1–100001–14 (2010).
6. Shiga, H. *et al.* *IEEE J. Solid-state Circuits* **45**, 142–152 (2010).
7. Das, S. & Appenzeller, J. *Nano Lett.* **11**, 4003–4007 (2011).
8. Chanthbouala, A. *et al.* *Nature Nanotechnol.* **7**, 101–104 (2011).
9. Choi, T., Lee, S., Choi, Y. J., Kiryukhin, V. & Cheong, S.-W. *Science* **324**, 63–66 (2009).

GENETICS

Broken giant linked to heart failure

Genetic mutations can cause a type of heart disease called dilated cardiomyopathy, by predisposing the organ to enlarge and function poorly. It has now been found that 27% of cases are due to mutations that disrupt the muscle protein titin.

ELIZABETH M. McNALLY

Cardiomyopathy — a disorder of heart muscle — is a leading cause of congestive heart failure, in which the organ cannot pump enough blood to meet the body's demands. Mutations account for many cardiomyopathy cases, but most of these gene variants are rare — often described from an individual or a single family¹. In particular, mutations associated with a form of the disease known as dilated cardiomyopathy have been identified in about 50 different genes, and the gene most commonly found to be mutated until now (*LMNA*) explains only 5% of cases. In a landmark study published in the *New England Journal of Medicine*, Herman *et al.*² report that as many as 27% cases, tested using modern sequencing techniques, can be attributed to dominant mutations in the *TTN* gene, which encodes titin, the largest human protein.

Titin is found in the sarcomeres, the contractile protein structures in heart muscle cells (Fig. 1). The protein is considered to be a molecular ruler that monitors the length of the sarcomere during muscle contraction and relaxation, and transmits this information to the rest of the cell. Consistent with this role, the activity of a carboxy-terminal kinase domain in titin is stimulated by muscle activity and stretch³.

Mutations in *TTN* have previously been linked⁴ to some cases of dilated cardiomyopathy (DCM), but comprehensive analyses of the gene and its mutations have been hampered by its complexity and exceptionally large size. *TTN* contains nearly 350 exons

(protein-coding regions), which encode a protein of about 35,000 amino acids, with many repeated sequence motifs. The arrival of faster and cheaper sequencing technologies now makes such analyses feasible.

To explore the link between *TTN* and DCM, Herman and colleagues² used an 'exon-capture' technique, in which they isolated and sequenced *TTN* exons from 312 patients with DCM, 231 patients with hypertrophic cardiomyopathy (HCM) and 249 control subjects without cardiomyopathy. The authors used two approaches for gene sequencing: massively

parallel (or 'next-generation') sequencing, and traditional Sanger sequencing. They found that 17% of the DCM patients had mutations that are likely to produce truncated forms of titin. This frequency increased to 27% for those DCM patients whose DNA was evaluated using massively parallel sequencing. By contrast, only about 1% of HCM cases and 3% of the control subjects had protein-truncating mutations. What's more, *TTN* mutations and DCM were co-inherited in those individuals whose families participated in the study. Taken together, these results clearly support the idea that mutations in *TTN* have a key role in DCM.

The *TTN* variants observed by the authors were not randomly distributed along the length of the gene: they were more frequent in an area coding for a titin region that interacts with the protein myosin in the sarcomere (Fig. 1). In particular, many of these mutations would disrupt titin's kinase domain, probably impairing the ability of the protein to signal to the cell that the sarcomere has lengthened abnormally⁵. The inability to sense or signal when the sarcomere has elongated too far may render heart cells

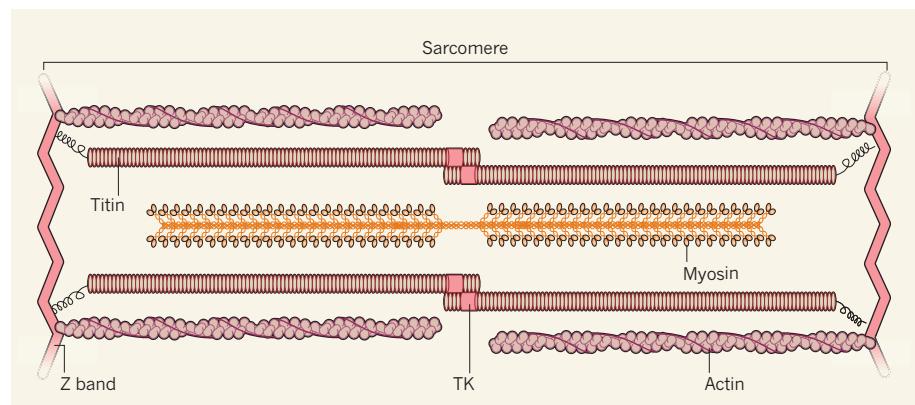


Figure 1 | A molecular ruler. Muscle fibres and heart cells are composed of repeated contractile units called sarcomeres that are made up of actin and myosin protein filaments, together with other proteins such as titin. Actin filaments are anchored to non-contractile protein structures at the borders of the sarcomere, known as 'Z bands'. Titin spans half the length of the sarcomere and is thought to act as a molecular ruler that senses changes in sarcomere length during muscle contraction and elongation. Herman *et al.*² report that 27% of patients with dilated cardiomyopathy, a disorder of heart muscle, have protein-truncating mutations in the titin-encoding gene. Many of these mutations would lead to disruption of a protein kinase domain (TK) in the protein, which would in turn result in an impaired ability to monitor sarcomere lengthening and, hence, in a malfunction of heart muscle.

susceptible to damage from muscle activity. Some of these mutations may also disrupt titin's ability to recoil during muscle contraction⁶.

In their analysis, Herman *et al.*² considered only rare, protein-truncating *TTN* variants (occurring with a frequency of less than 1%). So, how these mutations interact with the more common, non-truncating *TTN* mutations — such as those that change a single amino acid — or with cardiomyopathy-associated variants of other genes remains to be evaluated. Modelling the cumulative effect of multiple gene variants on the heart's physiology will require more sophisticated analyses.

The authors' data suggest that massively parallel sequencing is more sensitive than Sanger sequencing in detecting *TTN* variants. A reason for this could be that this type of sequencing determines a given sequence many more times than the older technique. This repeated sequencing, or possibly the analyses used to align the resulting sequences, may be more suited to the repetitive nature of *TTN*. In the immediate future, DNA microarrays designed to capture specific exons on the basis of bioinformatic identification⁷ could be used to analyse *TTN* variants. However, such microarrays should be carefully evaluated for their use in DCM diagnostics, as the boundaries of some of the *TTN* exons are incompletely defined in the most commonly used databases (those of the National Center for Biological Information in Bethesda, Maryland, and the University of California, Santa Cruz). Alternative approaches involving whole-genome sequencing might therefore be favoured for DCM diagnostics.

The actual outcome of DCM-related mutations, however, depends on the rest of an individual's genetic make-up and on environmental factors. For example, the effects of protein-disrupting *TTN* variants would be expected to be worse when combined with factors such as dysfunctional heart valves or hypertension. Interestingly, in the present study², men with *TTN* mutations had more severe DCM than women. Male hearts may be more sensitive to *TTN* mutations because they are normally larger than female hearts, and a larger heart faces greater strain and pressure.

Herman and colleagues' observation that a few of the control individuals carry titin-disrupting gene variants indicates that there are likely to be additional genetic or environmental factors that enhance the disease-causing potential of these mutations. Alternatively, these symptomless individuals might carry other gene variants that actively suppress the effects of the *TTN* mutations. For example, gene variants that drive increased degradation of the mutated *TTN* messenger RNA, or of the truncated titin, could reduce the amount of the damaged protein in the heart, and thereby limit the negative effects of *TTN* mutations. So, for a small percentage of people, life with less titin may not be too bad.

Diagnosing the genetic basis of DCM provides useful information that could guide therapy, especially when there is a risk of life-threatening — but treatable — irregular heart rhythms⁸. Most diagnostic testing now relies on sequencing many genes at once, as there are few clinical clues that reduce the number of candidate genes. However, given the frequency of titin-disrupting variation reported by Herman and colleagues, mutations in *TTN* now emerge as primary causes of, or highly potent contributors to, cardiomyopathy. ■

Elizabeth M. McNally is in the Department of Medicine, Section of Cardiology, and in the Department of Human Genetics, University of

Chicago, Chicago, Illinois 60637, USA.
e-mail: emcnally@uchicago.edu

1. Dellefave, L. & McNally, E. M. *Curr. Opin. Cardiol.* **25**, 198–204 (2010).
2. Herman, D. S. *et al.* *N. Engl. J. Med.* **366**, 619–628 (2012).
3. Miller, M. K., Granzier, H., Ehler, E. & Gregorio, C. C. *Trends Cell Biol.* **14**, 119–126 (2004).
4. Gerull, B. *et al.* *Nature Genet.* **30**, 201–204 (2002).
5. Gautel, M. *Pflügers Arch.* **462**, 119–134 (2011).
6. Nagueh, S. F. *et al.* *Circulation* **110**, 155–162 (2004).
7. Bamshad, M. J. *et al.* *Nature Rev. Genet.* **12**, 745–755 (2011).
8. van Rijssingen, I. A. *et al.* *J. Am. Coll. Cardiol.* **59**, 493–500 (2012).

The author declares competing financial interests. See go.nature.com/6gvod1 for details.

CONDENSED-MATTER PHYSICS

A duo of graphene mimics

The synthesis of analogues of graphene by two different means provides insight into the origins of massless particles and paves the way for studies of materials with exotic topological properties. SEE LETTERS P.302 & P.306

JONATHAN SIMON & MARKUS GREINER

When a particle is constrained to move in a honeycomb-lattice structure, its properties change dramatically: it behaves as though it has no mass and travels at the speed of light. Such particles, known as massless Dirac fermions, were first observed^{1,2} in atomic graphene in work that spawned a frenzy of research into its properties, both to understand the fundamental science and for technological applications. In a pair of letters published in this issue, Gomes *et al.*³ (page 306) and Tarruell *et al.*⁴ (page 302) describe how, for the first time, they have created synthetic analogues of graphene in two different systems.

Atomic graphene is a single layer of carbon atoms organized into a honeycomb structure. The new graphene analogues^{3,4} share this honeycomb topology and offer substantial advantages in manipulation and readout of material properties over what is possible with atomic graphene. Gomes *et al.*³ assembled 'molecular graphene' from individually placed carbon monoxide (CO) molecules, and studied the bizarre effects that arise from small variations in the material's lattice structure (Fig. 1a,b). Tarruell *et al.*⁴ show that a clever arrangement of laser beams (an optical lattice) produces a honeycomb structure to confine an ultracold gas of potassium atoms, and investigated how the lattice structure controls the mass of the atoms (Fig. 1c,d).

The first step in studying a graphene analogue is to show that its particles are massless. In everyday life, an object's mass is measured by weighing it on a scale. That is impossible for an extremely light object such as an electron, particularly when its mass is inextricably tied to the material in which it resides. Gomes and colleagues³ circumvented this difficulty by studying the energy gap between the valence and conduction energy bands in their molecular graphene. This quantity reflects the energy required to create an electron–hole pair out of the vacuum (where a hole is a charge carrier created by the absence of an electron), and thus the electron's mass. The authors measured the gap using scanning tunnelling microscopy and spectroscopy, and observed no separation between the valence and conduction bands — only a sharp dip at the Dirac point, at which the bands touch. The dip indicates a vanishing number of quantum states at zero energy, and the absence of a gap is a clear signature of massless Dirac fermions.

Having established the existence of massless electrons in their molecular graphene, Gomes and colleagues employed their exquisite control of the material to study how slight modifications to the lattice geometry change the properties of its resident electrons. First, they introduced a periodic arrangement known as a Kekulé structure to imbue their massless Dirac fermions with mass, and detected this mass through the appearance of an energy

gap in the material's electron-energy spectrum. They then showed that a distortion of the lattice structure, akin to squeezing it along several axes, makes the electrons act as if they are in a magnetic field⁵. The appearance of an isolated, zero-energy quantum state in the presence of this apparent field serves as striking confirmation of the existence of massless Dirac fermions in the molecular graphene.

Meanwhile, Tarruell and colleagues⁴ demonstrated that the potassium atoms in their material behave as massless Dirac fermions by watching them undergo a transition between the conduction and valence bands. Such transitions require substantial energy input for particles that have mass, to overcome the particles' rest mass and bridge the energy gap. But for a massless particle there is no energy gap, so the particle can move freely between valence and conduction bands at the Dirac point.

Holding a massless Dirac fermion near a Dirac point in order to observe transitions between bands would be extremely challenging in the solid state, because motional damping is difficult to control, and scattering off lattice impurities and phonons (quasi-particles associated with lattice vibrations) regularly randomizes the Dirac fermions' momentum. Ultracold atomic gases in optical lattices are ideally suited to overcoming this problem, as the atoms move without dissipation in extremely clean and tightly controlled environments⁶.

Tarruell and co-workers took advantage of these features in their work. They started with a cloud of fermionic (half-integer spin) potassium atoms in the valence band. These atoms exhibited a spread in momentum due to the Pauli exclusion principle, according to which two or more fermionic particles cannot occupy the same quantum state. The authors then subjected the atomic cloud to a magnetic-field gradient that gently accelerated the atoms, slowly varying their momentum. In this setting, any atom whose momentum trajectory crosses a Dirac point should be transferred to the conduction band (Fig. 1d).

Tarruell *et al.* elegantly observed the existence of Dirac points by imaging the momentum distribution of their cloud. They achieved this by allowing the cloud to expand and then taking an image of its spatial distribution, noting that atoms in the cloud that have more momentum move more quickly in this expansion. Before the acceleration, the atoms are distributed smoothly around zero momentum, whereas after acceleration the distribution has gaps corresponding to atoms that have been transported through a Dirac point, and hence transferred to the conduction band.

To address the question of how the Dirac fermions arise, Tarruell and colleagues smoothly tuned the energy potential of their structure to transform the underlying honeycomb lattice, which has two Dirac points, into a dimer lattice, which has none. To probe for

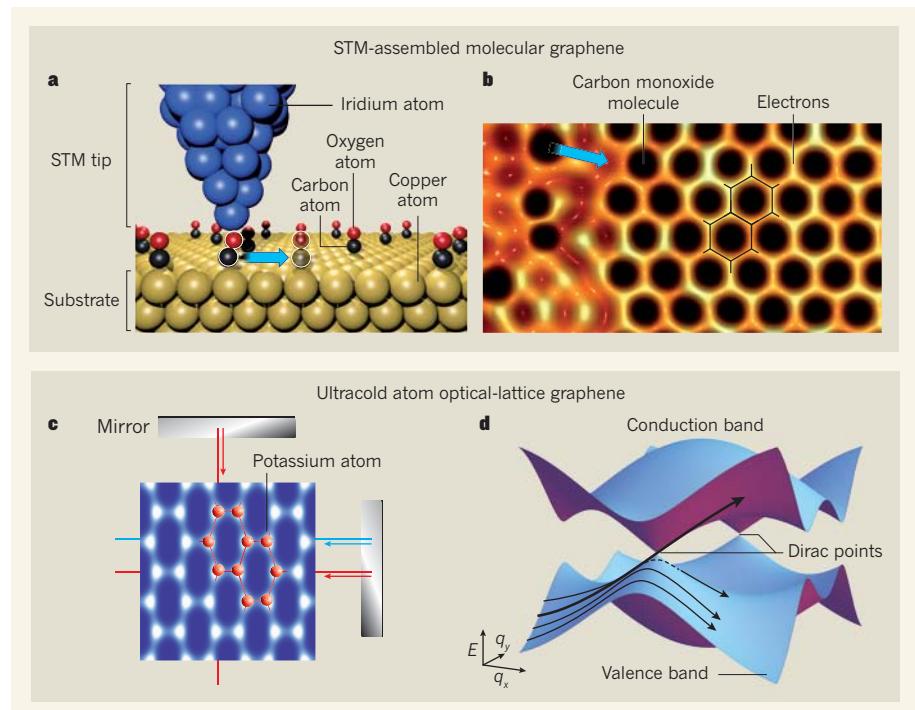


Figure 1 | Two approaches to making graphene analogues. **a**, Gomes *et al.*³ assembled molecular graphene using a scanning tunnelling microscope (STM), the tip of which is made of iridium atoms, to individually position (blue arrow) carbon monoxide (CO) molecules in a hexagonal pattern on a copper substrate. **b**, STM image of the partially assembled graphene. Electrons are repelled from the hexagonally arranged CO molecules and forced to move on a honeycomb grid (black lines). The blue arrow indicates the path of a CO molecule as it is positioned. **c**, Tarruell *et al.*⁴ generated a honeycomb optical lattice for confining ultracold potassium atoms through a combination of two interfering laser beams (red) and a third separate beam (blue), all retro-reflected by mirrors. The resulting intensity pattern is shown overlaid, with the intensity maxima shown in white and the tunnelling pathways between maxima emphasized by red lines. **d**, To observe the existence of Dirac fermions, the authors⁴ accelerated potassium atoms in the material's valence energy band and found that those that pass through a Dirac point are transferred to the conduction band. The plot displays the electron energy (E) as a function of its momenta in the x and y directions (q_x and q_y) for both bands. The arrows denote the paths of the potassium atoms through momentum space.

the existence of Dirac points in each lattice, they accelerated the potassium atoms and measured the fraction of atoms transferred to the conduction band. What they find is an impressive verification of a theoretically predicted topological phase transition^{7–9}: the atom fraction transferred drops abruptly to zero at precisely the point at which the dimer lattice forms and the two Dirac points merge and annihilate one another.

These studies^{3,4} pave the way for a new realm of condensed-matter physics, in which materials with exotic topological properties can be built to order from the ground up. However, some of the most interesting properties arise when interactions between particles cause them to self-organize into intricate patterns determined by the laws of quantum mechanics. To this end, it will be essential to suppress the role of the copper substrate underlying Gomes and colleagues' material, because it screens out the repulsive Coulomb interaction between the electrons and limits the electrons' lifetime.

By contrast, the ultracold atoms described by Tarruell *et al.*⁴ are extremely long lived and can exhibit strong interactions, but so far have

been studied only in synthetic magnetic fields¹⁰ that are not strong enough to investigate the physics of highly correlated Dirac fermions. Much stronger fields should be achievable through temporal modulation of the Dirac points¹¹. Once these limitations are overcome, the techniques developed by Tarruell, Gomes and their colleagues should point the way to a new generation of quantum materials. ■

Jonathan Simon and Markus Greiner
are in the Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mails: simon@physics.harvard.edu; greiner@physics.harvard.edu

1. Novoselov, K. S. *et al.* *Nature* **438**, 197–200 (2005).
2. Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. *Nature* **438**, 201–204 (2005).
3. Gomes, K. K., Mar, W., Ko, W., Guinea, F. & Manoharan, H. C. *Nature* **483**, 306–310 (2012).
4. Tarruell, L., Greif, D., Uehlinger, T., Jotzu, G. & Esslinger, T. *Nature* **483**, 302–305 (2012).
5. Levy, N. *et al.* *Science* **329**, 544–547 (2010).
6. Bloch, I., Dalibard, J. & Zwerger, W. *Rev. Mod. Phys.* **80**, 885–964 (2008).
7. Hasegawa, Y., Konno, R., Nakano, H. &

- Kohmoto, M. *Phys Rev B* **74**, 033413 (2006).
 8. Wunsch, B., Guinea, F. & Sols, F. *New J. Phys.* **10**, 103027 (2008).
 9. Montambaux, G., Piéchon, F., Fuchs, J.-N. & Goerbig, M. O. *Phys. Rev. B* **80**, 153412 (2009).
 10. Lin, Y.-J., Compton, R. L., Jiménez-García, K., Porto, J. V. & Spielman, I. B. *Nature* **462**, 628–632 (2009).
 11. Kitagawa, T., Berg, E., Rudner, M. & Demler, E. *Phys. Rev. B* **82**, 235114 (2010).

NEUROSCIENCE

How brains learn to control machines

After training, animals and humans can make their thoughts interact directly with computers. A study provides evidence that the corticostriatal system of the brain is essential for this learning process. **SEE LETTER P.331**

DAVID T. BLAKE

Brain-machine interfaces have a rich history in the sci-fi genre: in *The Matrix* films, human brains are plugged into a computer-based simulation that then becomes their 'reality'. But using our thoughts to directly control computers or other devices is not just in the realm of fantasy. Monkeys can learn to use visual cues to instruct a brain-machine interface to move a robotic arm or a computer cursor^{1,2}. And electrode arrays were implanted into the brain of a paralysed man in 2006, enabling him to control an artificial arm, to move a cursor on a computer screen and even to open e-mail³. Over time, an individual learns to improve their control over the brain-machine interface by modifying the activity of their brain, but how this happens is not well understood. On page 331 of this issue, Koralek *et al.*⁴ report that the corticostriatal system of the brain is involved in learning mental actions and skills that do not involve physical movement, such as those required for control of brain-machine interfaces*.

The corticostriatal system has a unique pattern of connectivity that enables sensory inputs to be associated with appropriate motor or cognitive responses⁵. It consists of a cortical component, the primary motor cortex, that exerts control over muscles, and a striatal component, the basal ganglia, that receives direct inputs from the motor cortex. The basal ganglia are involved in a wide range of learning conditions and are crucial to the motor deficits observed in Parkinson's and Huntington's diseases. Both corticostriatal components have a role in the learning and execution of physical skills requiring movement.

It was known that the learning of abstract (non-physical) skills, such as controlling a brain-machine interface, required the motor cortex, but it was not clear whether the striatal component was also involved. To address this

issue, Koralek and colleagues⁴ implanted electrodes in the cortex and striatum of live rats to record neuronal activity. The authors then set up a brain-machine interface (Fig. 1) in which a tonal sound with a specific pitch was automatically played to the animal depending on the measured activity of two groups of cortical neurons. The pitch of the sound increased if the first group of neurons was more active than the second group, and decreased if the second group was the more active. If the pitch got high enough, the rat earned a sugar-water reward; if the pitch dropped enough, then the animal received a food pellet.

Over a two-week period, the rats achieved a high level of proficiency at earning each reward. The researchers carried out a similar experiment with another set of animals, with the difference that no sound was played. In this case, the rats did not learn to alter activation of the two groups of neurons to earn rewards,

which confirmed the necessity of the auditory feedback. The animals that did learn developed a synchrony between the activity of their motor cortex and that of their striatum during task performance that was absent before learning. In a sense, the task learning caused the cortex and striatum to work in harmony. This synchrony is perhaps not entirely surprising, as it is known that the striatum and cortex are active and display similar temporal patterns of activity in a wide range of motor and cognitive procedures⁶.

Were the rats controlling the brain-machine interface in an intentional and goal-directed fashion? In a human brain-machine interface, we would simply ask the subject if they had a conscious knowledge that they were controlling an external interface. In the case of laboratory animals, such direct feedback is not possible, but Koralek *et al.* got part of the answer through clever experimental design. The authors gave sugar water to the rats before training, so that the animals were only motivated to get the food pellet afterwards. However, if the task performance were habitual and unintentional, the rats would not know how to earn one specific reward and not the other. What happened was that the animals did learn how to get the food pellet and not the sugar water, and this was associated with the change in brain activity predicted for goal-directed, intentional behaviour.

To assess the necessity for the striatum in the learning task, Koralek *et al.*⁴ repeated their experiments using genetically modified mice⁷ in which NMDA receptors were specifically inactivated in the striatum. NMDA receptors are required for normal functioning of the cortex–striatum connections, and so the genetically modified mice should have a greatly reduced ability to coordinate the two brain regions. Indeed, the authors observed that the mice could not learn how to achieve

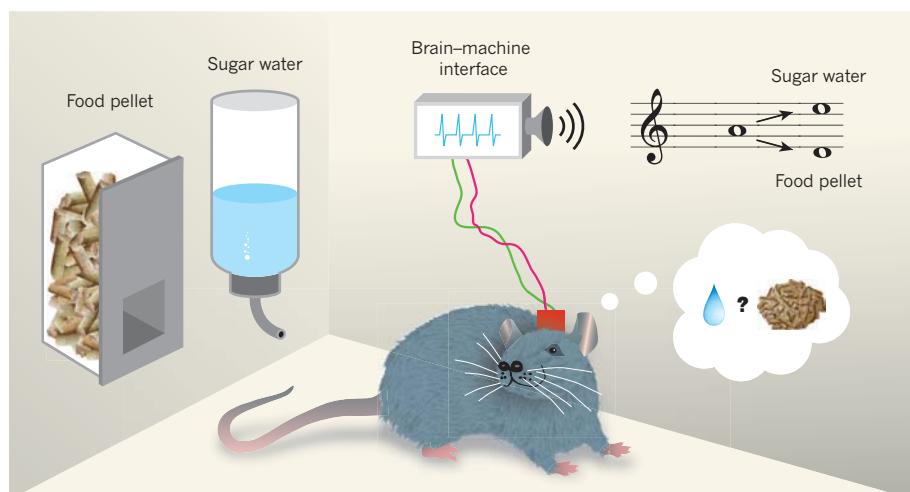


Figure 1 | Music to my brain. Koralek *et al.*⁴ implanted electrodes into the brains of live rats to record the neuronal activity of the motor cortex and the striatum. Depending on certain features of motor-cortex activity, a tonal sound with a specific pitch was automatically generated. The rats learnt to make the pitch of the sound rise or fall by modifying their brain activity, as they were rewarded with either sugar water or food pellets, respectively, if the pitch changed enough. By using genetically modified mice in similar experiments, the authors showed that activity of the striatum was required for the animals to learn the task.

*This article and the paper⁴ under discussion were published online on 4 March 2012.

the rewards, which supports the hypothesis that developing coordination between the striatum and the cortex is necessary for abstract learning.

Mechanistic studies of learning are inherently difficult. Most experiments sample activity before and after some type of learning and make inferences about what has happened in the interim. Others focus on certain tasks that can be learned in less than an hour or so, because this is the time frame during which stable recordings from single neurons can be reliably obtained. Implant technologies such as brain-machine interfaces have opened up new possibilities, because they allow researchers to record neuronal activity from a particular brain location over long periods of time. By resampling these locations in rats and mice throughout learning, Koralek *et al.* document that the animals can learn abstract skills by controlling neuronal

activity at multiple, specific, predetermined locations in the cerebral cortex, and that the neural basis for this learning also occurs in the cortex–striatum interplay. This finding augments the literature that inseparably ties motor functions and complex cognitive functions to the interconnected networks associated with the brain's frontal lobe.

Half a century ago, no one would have believed that it would be possible to implant a cardiac pacemaker into a person's chest that would restart their heart automatically. Today, these electrophysiological devices are common, as are cochlear implants for improved hearing, and as will soon be the case for retinal implants that will give at least part of their lost sight back to some blind people^{8,9}. Preliminary trials with brain-machine interfaces in humans are now in progress³, but are hampered by the limited stability of the primary link between the brain and the recording electrode. If the

link can be stabilized for longer durations, these interfaces may one day become as common as other electrophysiological devices and provide us with similarly useful services. ■

David T. Blake is at the Brain and Behavior Discovery Institute and in the Department of Neurology, Georgia Health Sciences University, Augusta, Georgia 30912, USA.
e-mail: dblake@georgiahealth.edu

1. Seruya, M. D. *et al.* *Nature* **416**, 141–142 (2002).
2. Carmena, J. M. *et al.* *PLoS Biol.* **1**, e42 (2003).
3. Hochberg, L. R. *et al.* *Nature* **442**, 164–171 (2006).
4. Koralek, A. C., Jin, X., Long, J. D. II, Costa, R. M. & Carmena, J. M. *Nature* **483**, 331–335 (2012).
5. Graybiel, A. M. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
6. Alexander, G. E., DeLong, M. R. & Strick, P. L. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
7. Tsien, J. Z. *et al.* *Cell* **87**, 1317–1326 (1996).
8. Humayun, M. S. *et al.* *Vision Res.* **43**, 2573–2581 (2003).
9. Wilke, R. *et al.* *Invest. Ophthalmol. Vis. Sci.* **52**, 5995–6003 (2011).

BIOCHEMISTRY

Favouring the unfavoured

The naturally occurring antibiotic lasalocid A contains a chemical structure that is not expected to form readily. The enzyme that catalyses the formation of this structure has been identified, and its activity is a revelation. SEE LETTER P.355

DAVID E. CANE

Lasalocid A and other polyether antibiotics are widely used to control parasitic infections in farm animals, and to enhance feed efficiency in cattle and other ruminants¹. They act by binding and carrying alkali-metal cations and protons across bacterial membranes, thereby interfering with essential ion gradients. The polyethers have also attracted considerable attention from scientists who want to identify the biosynthetic origins of these structurally complex substances and unravel the intricate pathways by which they are assembled². On page 355 of this issue, Hotta *et al.*³ report the intriguing mechanism of a step in the biosynthesis of lasalocid A, which they worked out using a combination of synthetic organic chemistry, enzymology, protein structural biology and quantum mechanics. Their findings provide insight into the molecular basis for the products formed by this fascinating class of enzymes.

Polyether molecules are characterized by a complex array of two or more cyclic ethers — oxygen-containing rings that, in polyether antibiotics, are typically composed of either five or six atoms. The carbon backbones of polyethers are assembled by enzymes known

as polyketide synthases from simple building blocks such as acetate, propionate and butyrate, which contain two, three and four carbon atoms, respectively.

A general model for the biological formation of polyethers was formulated⁴ in 1983. It postulates that intermediates called epoxides (cyclic ethers consisting of three atoms) undergo a cascade of enzyme-catalysed reactions to form the polyether rings. The energetics for this process resembles that of the gravity-driven tumbling of a line of dominoes: the enzyme triggers the reaction cascade, just as the fall of the first domino causes the other dominoes to topple. In the case of polyethers, however, the enzyme also controls the number and size of the ether rings that form, which is analogous to steering the tumbling dominoes to control precisely where each one lands.

In the past decade, considerable progress has been made in elucidating the steps of polyether biosynthesis. The enabling advance was the identification^{2,5–8} of bacterial gene clusters that control the formation of several important polyethers. Each gene in these clusters encodes an enzyme that catalyses a specific step in the biosynthetic pathway. Selectively knocking out some of the genes, to interrupt the pathway when the remaining genes in the cluster

are expressed in cells, has enabled the functions of the deleted genes to be established. As a result of such deletions, several compounds that accumulate have been identified as key intermediates in the pathway, including the unsaturated polyketides that are initially generated from the acetate, propionate and butyrate building blocks.

Similar studies have also begun to clarify the final stages of polyether biosynthesis. For example, oxidative enzymes that generate the key epoxide intermediates (known as poly-epoxides) from unsaturated polyketides have been discovered, as have the epoxide hydrolase enzymes that catalyse the reaction cascade in which the polyepoxides are converted into polyethers². A group led by Hideaki Oikawa (one of the authors of the new paper³) recently expressed and purified⁹ an intriguing epoxide hydrolase, Lsd19. The team established that this enzyme is responsible for converting a precursor molecule harbouring two epoxide rings into a mixture of two closely related polyethers — lasalocid A and isolasalocid A (Fig. 1). Both compounds contain two cyclic ethers, but the second of these in lasalocid A is a six-membered ring, whereas in isolasalocid A it is a five-membered ring.

The authors went on to show⁹ that Lsd19 consists of two halves — Lsd19A and Lsd19B — that have similar amino-acid sequences. They found that Lsd19A catalyses the formation of the first ether and Lsd19B controls the formation of the second. The team also identified acidic and basic amino-acid residues at the active sites of each domain, and reported evidence implicating these residues in the catalysis of the reactions that convert their epoxide substrates to cyclic ethers.

Hotta *et al.*³ now report the high-resolution (1.59-ångström) X-ray crystal structure of Lsd19, in which a synthetic, uncyclized substrate analogue is bound to the half of the

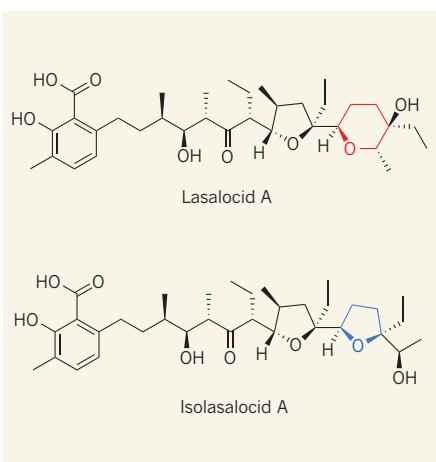


Figure 1 | Polyether antibiotics. The naturally occurring antibiotic lasalocid A is biosynthesized as a mixture with its isomer, isolasalocid A. The structures differ in that the terminal cyclic ether (red) in lasalocid A consists of six atoms, whereas the equivalent ether in isolasalocid A (blue) has five. Hotta *et al.*³ report the mechanism of formation of the six-membered ether in lasalocid A.

protein corresponding to Lsd19A and a synthetic product analogue carrying two ether rings is bound to the active site of the Lsd19B domain. The structure clearly shows that the active-site cavities of each half of the protein have very similar overall architectures and general catalytic features, but different depths. The consequent differences in active-site shape and volume can reasonably account for the distinct role of each domain in catalysing different stages of the two-step, ether-forming reaction cascade. The high-resolution structure also reveals candidates for the catalytically important acidic and basic residues in the active sites of each protein subunit.

The authors then probed the enzyme in even more depth by carrying out theoretical studies to model the reactions catalysed by Lsd19B, using quantum-mechanical calculations that incorporated the experimentally determined positions of the catalytic residues in the active site of the Lsd19B domain. In particular, they wanted to learn how the enzyme preferentially catalyses the formation of a six-membered cyclic ether, given that a five-membered ring would normally form more easily in the non-enzyme-catalysed reaction (according to well-established rules¹⁰ for predicting the outcomes of ring-formation reactions in solution).

Hotta *et al.* began by modelling ether-ring formation in the presence of either a simple acid or a base, but in the absence of an enzyme. As expected, the theoretical model predicted that the five-membered ring forms more quickly than the six-membered ring, although the preference is smaller in the case of the base-catalysed reaction. Despite this difference in formation rate, the authors concluded that the product containing the

six-membered ring is thermodynamically more stable than that containing the five-membered ring, even though it is more difficult to obtain (slower to form).

By contrast, when they modelled ether-ring formation in the presence of the enzyme, the calculations showed that formation of the six-membered ether was favoured. This is a chemical process in which an aspartate amino acid acts as a catalytic base, and two other amino-acid residues act as acids to stabilize the transition state of the reaction. The results show that when enzymes have the choice of catalysing two competing chemical reactions, they can evolve to preferentially accelerate the intrinsically slower transformation, until the otherwise less-preferred product becomes dominant. In other words, enzyme catalysts can channel chemical ‘dominoes’ to fall in a different direction from that favoured in the absence of the enzyme.

This impressive study solidifies and extends earlier hard-won advances in the study of polyether biosynthesis, and has implications for the formation of many other polyethers. Still unanswered is how the latest findings might apply to the biosynthesis of polyethers that have more than two ether rings, and to

the intriguing ladder-like polyethers found in toxins made by marine organisms. A complete picture of polyether biosynthesis will also require a better understanding of how the oxidative enzymes involved in earlier stages of polyether biosynthesis control the three-dimensional structure of the polyepoxides that serve as substrates for the intricate cascade of ether-forming reactions. ■

David E. Cane is in the Department of Chemistry, Brown University, Providence, Rhode Island 02912-9108, USA.
e-mail: david_cane@brown.edu

1. Callaway, T. R. *et al.* *Curr. Issues Intest. Microbiol.* **4**, 43–51 (2003).
2. Liu, T., Cane, D. E. & Deng, Z. *Meth. Enzymol.* **459**, 187–214 (2009).
3. Hotta, K. *et al.* *Nature* **483**, 355–358 (2012).
4. Cane, D. E., Celmer, W. D. & Westley, J. W. *J. Am. Chem. Soc.* **105**, 3594–3600 (1983).
5. Oliynyk, M. *et al.* *Mol. Microbiol.* **49**, 1179–1190 (2003).
6. Sun, Y. *et al.* *Chem. Biol.* **10**, 431–441 (2003).
7. Smith, L., Hong, H., Spencer, J. B. & Leadlay, P. F. *ChemBioChem* **9**, 2967–2975 (2008).
8. Migita, A. *et al.* *Biosci. Biotechnol. Biochem.* **73**, 169–176 (2009).
9. Minami, A. *et al.* *Org. Lett.* **13**, 1638–1641 (2011).
10. Baldwin, J. E. *J. Chem. Soc. Chem. Commun.* 734–736 (1976).

GENE EXPRESSION

Transcription initiation unwrapped

A genome-wide, high-resolution study of DNA-binding sites for proteins that transcribe DNA into RNA reveals details about how this process occurs *in vivo*.
SEE ARTICLE P.295

STEPHEN BURATOWSKI

Decades of research using purified molecules *in vitro* have produced a basic understanding of the enzymes and mechanisms that contribute to gene expression in eukaryotes (organisms such as animals, plants and fungi). But these results must be confirmed in living cells, and a productive approach has been to use chromatin immunoprecipitation. This technique reveals the genomic locations of DNA-binding proteins such as those forming nucleosomes — DNA segments wrapped around a histone protein core — and transcription factors. Although these studies have produced a wealth of data, they have not always provided mechanistic insight. However, a paper by Rhee and Pugh¹ on page 295 of this issue sheds light on several questions concerning the initiation of DNA's transcription into RNA.

In chromatin immunoprecipitation (ChIP),

cells are first treated with a chemical that crosslinks proteins and DNA. The cells are then disrupted so that their DNA is fragmented. By using specific antibodies, a protein of interest is isolated together with any bound DNA pieces, and these can then be analysed by high-throughput DNA sequencing or other techniques. This allows a quick identification of all the binding sites for a protein, such as those that form protein complexes for transcription initiation, across an entire genome.

Rhee and Pugh have previously reported² a modification of ChIP, called ChIP-exo, in which an enzyme removes all DNA except that closest to the protein–DNA crosslink, markedly improving the technique's resolution to a few DNA nucleotides. The results can be quite striking, as illustrated by the excellent correlation that the authors observe in their present study¹ between the ChIP-exo crosslink sites for the transcription factors TBP (TATA-binding protein) and TFIIB, and the protein–DNA

contacts seen in their crystal structures.

The researchers analyse promoters — sequences that specify where to begin the transcription of DNA into RNA — in cells of the yeast *Saccharomyces cerevisiae*, a model eukaryotic organism. Most notable are their findings regarding how the enzyme that synthesizes messenger RNA, called RNA polymerase (Pol II), is targeted to promoters. TBP is known to recognize the 'TATA box' — a specific DNA sequence found in many promoters — and to position Pol II and its associated factors at the transcription start site (TSS). However, only some promoters, typically those that alternate between repressed and highly active states, contain an obvious TATA box sequence, which represents the optimal TBP-binding site³.

Surprisingly, Rhee and Pugh's analysis¹ of TBP-binding sites in 'TATA-less' promoters — more prevalent among 'housekeeping' genes that are expressed ubiquitously — reveals this to be a misnomer. These promoters do contain TATA boxes, but their sequences stray from the standard sequence by two or more DNA bases and so their binding to TBP is weaker. This finding echoes those from classic studies⁴ on the yeast *HIS3* promoter, which contains two TATA boxes: a weak one for constitutive basal transcription and another, clearly recognizable, for maximal, regulated expression.

Although all Pol II promoters seem to share a common mode of binding to TBP and basal factors, Rhee and Pugh's data¹ help to explain a functional distinction that has been observed⁵ between the two classes of promoter. The authors report¹ that, at promoters with obvious TATA sequences, the TSS — and even the initiation complex itself — often overlaps with the first nucleosome in the transcribed region. The expression of genes containing such promoters tends to depend on the presence of the SAGA protein complex⁵, which facilitates nucleosome movement, and thus DNA unwrapping, by adding acetyl groups to the nucleosome's histones. Therefore, the first nucleosome probably represses the genes' transcription by blocking their TSS, and gene activation occurs when the histones are removed by targeted acetylation. Once the TSS-containing DNA is unwrapped, efficient binding of TBP, and Pol II and its associated factors, allows the promoter to be expressed at very high levels.

However, most genes have less-obvious TATA boxes, and their expression depends on other proteins known as TBP-associated factors (TAFs), which together with TBP constitute TFIID. *In vitro* studies⁶ have shown that TAFs interact with DNA sequences downstream of the TATA box, including sequences around the TSS. These additional contacts may help to compensate for the weaker TBP binding to the DNA, but they probably have other functions. Rhee and Pugh¹ find that, at those promoters to which TFIID preferentially binds, TSSs are located near the upstream

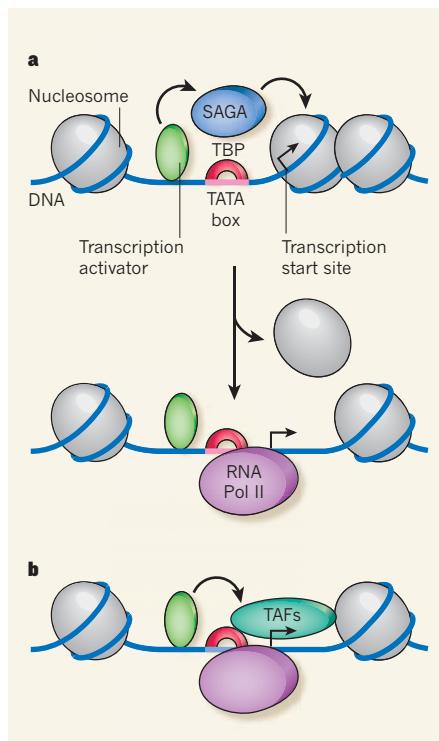


Figure 1 | Two ways of starting the synthesis of messenger RNA. Promoters are sequences within genes that specify where DNA's transcription into RNA starts. **a**, Some promoters have a clearly recognizable sequence, or TATA box, to which transcription factor TBP can bind to recruit the enzyme RNA polymerase (Pol II), which synthesizes messenger RNA. These promoters can alternate between repressed and active states. In the repressed state (top), a nucleosome (protein–DNA complex) blocks the transcription start site. Transcription activator proteins can then recruit additional proteins, such as the SAGA complex, to trigger nucleosome removal, allowing access to Pol II and therefore activating transcription (bottom). **b**, Most promoters lack a clearly recognizable TATA box, and their expression depends on the presence of transcription factor TFIID, a complex formed by TAF proteins and TBP. Transcription activators could recruit TAFs, which, in turn, might interact with both TBP and a nucleosome to keep the transcription start site accessible.

boundary of the first nucleosome. Therefore, TAFs may be positioned in such a way that they contact the first nucleosome, preventing it from encroaching on the promoter and thereby allowing basal gene expression. Indeed, some TAFs form a structure resembling the nucleosome histone core⁷, suggesting that they might slot into position within an array of nucleosomes.

In addition to invalidating the concept of TATA-less promoters, Rhee and Pugh raise questions about two other recent hypotheses. The first proposes that Pol II promoters are intrinsically bidirectional, that is, a single TATA box can drive transcription in opposite directions. This idea seems plausible because TATA sequences are roughly palindromic, and

transcript sequencing studies^{8,9} have shown that the TSSs of many mRNAs are close to a non-coding RNA that is transcribed in the opposite direction. However, the authors' ChIP-exo data¹ show that the nucleosome-depleted regions between these divergent TSSs harbour two initiation complexes. In other words, bidirectional transcription is the result of two overlapping but divergent promoters driving transcription in opposite directions, rather than a single promoter that can fire in both directions.

The second hypothesis¹⁰ is that 'gene looping' — the formation of a physical linkage between the beginning and end of active genes — is mediated, in part, by TFIIB. This model is based on observed interactions between mutations in genes that encode TFIIB and 3'-end processing factors (which modify the end of mRNA precursors), as well as ChIP localization of TFIIB (but not the rest of the initiation complex) at transcription-termination regions of selected genes in yeast¹⁰. However, the present study¹ and another genome-wide ChIP analysis¹¹ failed to localize TFIIB to 3' ends, except in the context of initiation complexes at an adjacent promoter. Therefore, the general role of TFIIB in gene looping needs further scrutiny.

It is worth noting that two transcription factors occupying the same genomic location in ChIP experiments may not actually be there at the same time in the same cell, as this technique captures a snapshot of events in a cell population. *In vitro* experiments are therefore needed to probe the kinetics and intermediates of gene expression. Rhee and Pugh¹ use many biochemical and structural studies to inform the interpretation of their ChIP-exo data; the ChIP-exo data, in turn, provide an essential *in vivo* test for *in vitro*-derived molecular models. This synergy underscores the necessity of applying both approaches to important questions in gene expression. ■

Stephen Buratowski is in the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: steveb@hms.harvard.edu

1. Rhee, H. S. & Pugh, B. F. *Nature* **483**, 295–301 (2012).
2. Rhee, H. S. & Pugh, B. F. *Cell* **147**, 1408–1419 (2011).
3. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. *Cell* **116**, 699–709 (2004).
4. Struhl, K. *Mol. Cell. Biol.* **6**, 3847–3853 (1986).
5. Huisinga, K. L. & Pugh, B. F. *Mol. Cell* **13**, 573–585 (2004).
6. Verrijzer, C. P., Chen, J. L., Yokomori, K. & Tjian, R. *Cell* **81**, 1115–1125 (1995).
7. Selleck, W. *et al.* *Nature Struct. Mol. Biol.* **8**, 695–700 (2001).
8. Neil, H. *et al.* *Nature* **457**, 1038–1042 (2009).
9. Xu, Z. *et al.* *Nature* **457**, 1033–1037 (2009).
10. Singh, B. N. & Hampsey, M. *Mol. Cell* **27**, 806–816 (2007).
11. Mayer, A. *et al.* *Nature Struct. Mol. Biol.* **17**, 1272–1278 (2010).

Ancient deuterostome origins of vertebrate brain signalling centres

Ariel M. Pani^{1,2}, Erin E. Mullarkey^{3*}, Jochanan Aronowicz^{4*}, Stavroula Assimacopoulos⁵, Elizabeth A. Grove^{3,4,5} & Christopher J. Lowe^{1,2,4}

Neuroectodermal signalling centres induce and pattern many novel vertebrate brain structures but are absent, or divergent, in invertebrate chordates. This has led to the idea that signalling-centre genetic programs were first assembled in stem vertebrates and potentially drove morphological innovations of the brain. However, this scenario presumes that extant cephalochordates accurately represent ancestral chordate characters, which has not been tested using close chordate outgroups. Here we report that genetic programs homologous to three vertebrate signalling centres—the anterior neural ridge, zona limitans intrathalamica and isthmic organizer—are present in the hemichordate *Saccoglossus kowalevskii*. *Fgf8/17/18* (a single gene homologous to vertebrate *Fgf8*, *Fgf17* and *Fgf18*), *sfrp1/5*, *hh* and *wnt1* are expressed in vertebrate-like arrangements in hemichordate ectoderm, and homologous genetic mechanisms regulate ectodermal patterning in both animals. We propose that these genetic programs were components of an unexpectedly complex, ancient genetic regulatory scaffold for deuterostome body patterning that degenerated in amphioxus and ascidians, but was retained to pattern divergent structures in hemichordates and vertebrates.

During vertebrate development, the brain arises from coarsely patterned planar neuroectoderm through successive refinement of regional identities, resulting after morphogenesis and growth in a complex structure composed of highly specialized areas^{1–3}. In contrast, invertebrate chordates have relatively simple nervous systems that lack unambiguous homologues of many vertebrate brain regions^{4–6}. These clear disparities in nervous system complexity indicate that key innovations in patterning mechanisms have attended the evolution of the vertebrate brain from a simpler central nervous system (CNS). However, attempts to identify the presumed genetic regulatory novelties have been mainly inconclusive, and comparative studies have mostly revealed similarities, rather than differences, in the early transcriptional architectures of diverse bilaterian nervous systems^{7–12}. Notable exceptions are CNS signalling centres, which have emerged as strong candidates for vertebrate genetic regulatory novelties involved in early vertebrate brain evolution^{8–10}. These centres act as secondary organizers that mediate regional patterning in the CNS and are often necessary and sufficient for the establishment of vertebrate-specific brain structures^{1–3,13–18}.

The anterior neural ridge (ANR), zona limitans intrathalamica (ZLI) and isthmic organizer (IsO) are the three primary signalling centres that direct anteroposterior patterning in the vertebrate anterior neural plate and then later in the developing brain^{1–3}. Homologous signalling centres and their molecular signatures are absent, or divergent, in amphioxus and ascidians^{8–10,19–25}, consistent with the idea that they are vertebrate novelties whose origins could have driven CNS innovations in stem vertebrates^{8–10}. However, this idea depends on amphioxus adequately representing ancestral states for chordate developmental genetic characters, and does not account for the possibility of secondary losses in cephalochordates. Here we present developmental data from the hemichordate *S. kowalevskii* to test an alternative scenario for signalling-centre origins; namely, that ANR, ZLI and IsO genetic programs pre-date chordate origins and

were secondarily simplified or lost along the lineages leading to the invertebrate chordates.

Hemichordates are a deuterostome phylum closely related to chordates²⁶ and are a promising outgroup for investigating chordate evolution. Previous studies established that despite substantial body-plan divergence between the two groups, hemichordates and vertebrates share a broadly conserved transcriptional regulatory architecture during early body patterning^{7,27–29}, which is demonstrated by the close similarities in spatial arrangements of expression domains of many transcription factors that are involved in early bilaterian anteroposterior patterning. This combination of morphological divergence and developmental genetic similarity makes hemichordates an informative outgroup for testing the proposed coupling of vertebrate morphological and developmental genetic innovations. Here, we present descriptive and functional evidence that genetic programs homologous to the ANR, ZLI and IsO are present in *S. kowalevskii*, indicating that they were elements of an ancient developmental genetic toolkit for deuterostome body patterning that were subsequently modified and elaborated in stem vertebrates to regulate brain development.

An ANR-like regulatory program in hemichordates

The ANR is located in the anterior neural plate of vertebrates and is a source of fibroblast growth factors (FGFs) and secreted frizzled-related proteins (SFRPs), which establish and pattern the telencephalon^{1,2,15,16,30,31}. Topologically consistent with vertebrate CNS expression domains, *sfrp1/5* and *fgf8/17/18* are expressed in *S. kowalevskii* anterior proboscis ectoderm (Fig. 1a–d). In vertebrates, FGFs and antagonists of the Wnt pathway mediate ANR function and telencephalon patterning^{15,16,30–32}. In mice, conditional *Fgfr1*, *Fgfr2* and *Fgfr3* knockout abolishes the telencephalon¹⁶, whereas *Fgf8* mutants have a posteriorized neocortex³³. Similarly, treating *S. kowalevskii* embryos with the FGF receptor inhibitor SU5402 (ref. 34) from late gastrula through to

¹Committee on Evolutionary Biology, The University of Chicago, 1025 East 57th Street, Chicago, Illinois 60637, USA. ²Hopkins Marine Station, Department of Biology, Stanford University, 120 Oceanview Boulevard, Pacific Grove, California 93950, USA. ³Committee on Neurobiology, The University of Chicago, 947 East 58th Street, Chicago, Illinois 60637, USA. ⁴Department of Organismal Biology and Anatomy, The University of Chicago, 1027 East 57th Street, Chicago, Illinois 60637, USA. ⁵Department of Neurobiology, The University of Chicago, 947 East 58th Street, Chicago, Illinois 60637, USA.

*These authors contributed equally to this work.

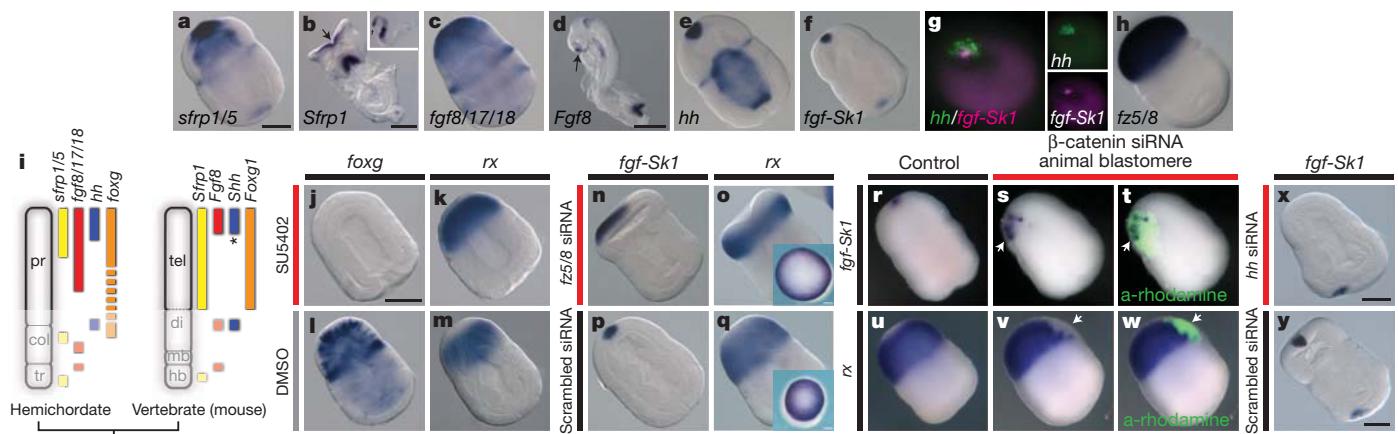


Figure 1 | An ANR-like signalling centre in *S. kowalevskii*. **a–h**, *S. kowalevskii* and mouse *in situ* hybridizations for markers of ANR and telencephalon. *S. kowalevskii* embryos are at the double-groove stage (36 h), and are shown in dorsal view with the anterior (proboscis) to the top left of the image, except where noted. Embryos are optically cleared except in **o**, **q** (insets) and **r–w**. Mouse embryos are at approximately embryonic day 8.5. **a**, *S. kowalevskii* *sfrp1/5* expression. **b**, Frontal view of mouse *Sfrp1* expression; arrowhead denotes ANR. Inset shows lateral view. **c**, *S. kowalevskii* *fgf8/17/18* expression. **d**, Mouse *Fgf8* expression; arrowhead denotes ANR. **e**, *S. kowalevskii* *hh* expression. **f**, *S. kowalevskii* *fgf-Sk1* expression. **g**, Frontal view of double FISH for *hh* and *fgf-Sk1*. **h**, *fz5/8* expression. **i**, Anteroposterior expression topologies in *S. kowalevskii* and mouse embryos. Anterior to top. Asterisk indicates *Shh* is expressed in the medial ganglionic eminence, near the ANR. **j–k**, *foxg* (**j**) and *rx* (**k**) expression in embryos treated with SU5402. **l**, **m**, *foxg* (**l**) and *rx*

double-groove stages (28–36 h) eliminated the morphological boundary between the proboscis and collar, and resulted in altered gene expression domains, indicating loss and/or posteriorization of anterior proboscis (Fig. 1j–m). Expression of *foxg*, the *S. kowalevskii* homologue of the telencephalon marker *foxg1*, was completely eliminated (Fig. 1j, l), whereas *rx*, which is normally excluded from the anterior proboscis, was expressed up to the anterior limit of the embryo (Fig. 1k, m).

Wnt antagonists expressed in the ANR and/or telencephalon are also critical for forebrain development in vertebrates^{2,15,35}. To explore potential similarities in Wnt functions between the telencephalon and proboscis, we first suppressed Wnt activity in the developing proboscis by using short interfering RNA (siRNA) microinjections to knock down the Wnt receptor *frizzled 5/8*, which is expressed with a sharp boundary at the proboscis base (Fig. 1h). *Fz5/8* siRNA expanded apical identity at the expense of the posterior proboscis, as demonstrated by expansion of the apical marker *fgf-Sk1* (Fig. 1n, p) and contraction of *rx* expression to a more posterior domain (Fig. 1o, q). Embryos that were injected with a scrambled control siRNA were indistinguishable from wild-type embryos (Fig. 1p, q), and the effects of *fz5/8* knockdown were limited to its endogenous expression domain (Supplementary Fig. 1). Second, we suppressed Wnt signalling in small patches of proboscis ectoderm by injecting β -catenin siRNA into single animal blastomeres at early cleavage stages. β -catenin-deficient clones in the posterior proboscis ectoderm expressed the apical marker *fgf-Sk1* (Fig. 1r–t), but not the more posterior proboscis marker *rx* (Fig. 1u–w), which indicates a transformation of posterior to apical fate in the absence of local Wnt signalling.

Although not a genetic component of the ANR itself, *sonic hedgehog* (*Shh*) is expressed in the nearby medial ganglionic eminence^{1,36} (see Fig. 2c, d) and interacts with FGFs to regulate telencephalon patterning³⁷. The apical *S. kowalevskii* proboscis ectoderm expresses *hedgehog* (*hh*; a homologue of *Shh*) (Fig. 1e) in a domain that partially overlaps *fgf-Sk1* (Fig. 1f), with *hh* more broadly expressed dorsally and laterally (Fig. 1g). Embryos injected with *hh* siRNA lacked apical *fgf-Sk1* expression (Fig. 1x, y) indicating that *hh* at least partially regulates anterior FGF signalling and apical–ventral patterning. This finding

(m) expression in embryos treated with DMSO. **n**, Expanded apical *fgf-Sk1* expression in an embryo injected with *fz5/8* siRNA. **o**, Retracted *rx* expression in an embryo injected with *fz5/8* siRNA. **p**, *fgf-Sk1* expression in a siRNA control embryo. **q**, *rx* expression in a siRNA control embryo. Insets in **o**, **q** show frontal views of uncleared embryos. **r**, Wild-type *fgf-Sk1* expression. **s**, *fgf-Sk1* expression in descendants of a blastomere injected with β -catenin siRNA. **t**, Merged darkfield and fluorescence images showing clonal descendants of the injected cell (green). **u**, Wild-type *rx* expression. **v**, *rx* is not expressed in descendants of a blastomere injected with β -catenin siRNA. **w**, Merged darkfield and fluorescence images showing the location of the β -catenin-deficient clone (green). **x**, *fgf-Sk1* expression in an embryo injected with *hh* siRNA. **y**, *fgf-Sk1* expression in a siRNA control embryo. Scale bars, 100 μ m in *S. kowalevskii*, and 200 μ m (**b**) and 500 μ m (**d**) in mice. col, collar; di, diencephalon; hb, hindbrain; mb, midbrain; pr, proboscis; tel, telencephalon; tr, trunk.

raises the possibility that an anterior signalling domain including FGFs, Wnt antagonists and *hh* was present in stem deuterostomes and was spatially partitioned during vertebrate evolution.

A ZLI-like regulatory program in hemichordates

The ZLI is located in the vertebrate diencephalon and specifies the flanking prethalamus and thalamus^{1,17,18,23}. The molecular signature of the ZLI is a narrow, transverse domain of *Shh* expression (Fig. 2c, d) that patterns the mid-diencephalon along its anteroposterior axis^{17,18,23}. A homologous *hh* expression domain is absent in invertebrate chordates, supporting the idea that the ZLI is a vertebrate genetic innovation, possibly associated with forebrain origins^{10,21–23}. Notably, we found that *S. kowalevskii* *hh* is expressed in a circumferential ectodermal band at the proboscis–collar boundary (Fig. 2a) in a transcriptional context of expression domains that is similar to the vertebrate forebrain⁷. Expression of the *Hh* receptor *ptch* in an overlapping but broader domain than *hh* itself, indicates that *hh* signals to adjacent ectoderm (Fig. 2b) similar to the ZLI (Fig. 2e, f). The proboscis–collar boundary region also expresses *fng*, *otx*, *wnt8* (Fig. 2g–m, v) and *six3* (ref. 7) in spatial arrangements that are characteristic of the mid-diencephalon and ZLI. In vertebrates, *lfng* is expressed broadly in the diencephalon except for the ZLI (ref. 38), whereas *Wnt8b* and *Otx* genes mark the ZLI itself in a combinatorial pattern that is unique to this region¹⁸ (Fig. 2i, j, m). *S. kowalevskii* *hh* is also expressed in a *fng*-negative, *otx*- and *wnt8*-positive territory posterior to *six3* (ref. 7) (Fig. 2g, h, k, l, v). Notably, the hemichordate homologues of *irx* and *fezf*, which abut at the ZLI and are expressed in anterior abutting domains in *Drosophila* and amphioxus (ref. 10), show divergent expression patterns in *S. kowalevskii* (Supplementary Fig. 2), indicating secondary modification in hemichordates. In vertebrates, *Otx* genes are expressed in restricted domains in the mid-diencephalon at the stage when the ZLI forms (Fig. 2i, j), and *otx1l* and *otx2* knockdown eliminates *shh* expression at the zebrafish ZLI³⁹. Similarly, in *S. kowalevskii*, *hh* and *otx* are co-expressed in the presumptive proboscis–collar boundary ectoderm shortly after gastrulation (Fig. 2k), and *otx* siRNA knockdown reduced *hh* expression at the proboscis–collar boundary (Fig. 2w, x).

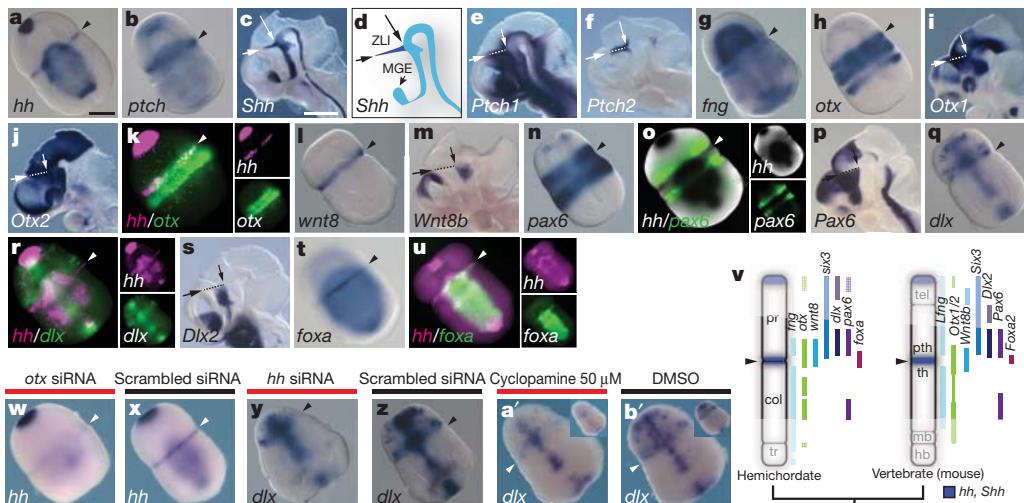


Figure 2 | A ZLI-like signalling centre in *S. kowalevskii*. **a–u**, *In situ* hybridizations for *S. kowalevskii* and mouse homologues of ZLI and diencephalic markers. Arrowheads mark the proboscis–collar boundary in *S. kowalevskii*. Mouse images show hemisected heads at embryonic day 10.5, and dashed lines indicate the ZLI, with arrows denoting its extent. **a**, *S. kowalevskii hh* expression. **b**, *S. kowalevskii ptc1* expression (see Supplementary Fig. 3). **c**, mouse *Shh* expression. **d**, Diagram of *Shh* expression showing ZLI in dark blue. **e**, Mouse *Ptc1* expression. **f**, Mouse *Ptc2* expression. **g**, *S. kowalevskii fng* expression. **h**, *S. kowalevskii otx* expression. **i**, Mouse *Otx1* expression. **j**, Mouse *Otx2* expression. **k**, *hh* (magenta) and *otx* (green) are co-expressed at the presumptive proboscis–collar boundary. **l**, *S. kowalevskii wnt8* expression. **m**, Mouse *Wnt8b* expression. **n**, *S. kowalevskii pax6* expression. **o**, Double *in situ* hybridization showing *pax6* expression (fluorescence, green) anterior to *hh*

(colorometric, black). **p**, Mouse *Pax6* expression. **q**, *S. kowalevskii dlx* expression. **r**, Double FISH showing *dlx* expression (green) in the proboscis base anterior to *hh* (magenta). **s**, Mouse *Dlx2* expression. **t**, *S. kowalevskii foxa* expression. **u**, Double FISH showing *foxa* (green) and *hh* (magenta) expression. **v**, Diagram of anteroposterior expression topologies of ZLI and forebrain marker homologues in *S. kowalevskii* and mice. *Six3* expression based on previous data⁷. Anterior to top. **w**, **x**, *otx* siRNA downregulates *hh* expression at the proboscis–collar boundary (**w**) relative to a control siRNA (**x**). **y**, **z**, *hh* siRNA reduces *dlx* expression in the proboscis base (**y**) relative to a scrambled siRNA (**z**). **a'**, **b'**, cyclopamine treatment reduces *dlx* expression in the proboscis base (**a'**) relative to a control embryo treated with DMSO (**b'**). Insets show ventral views. Scale bars, 100 μ m in *S. kowalevskii* embryos, 1 mm in mice. MGE, medial ganglionic eminence; pth, prethalamus; th, thalamus.

Hemichordate homologues of diencephalic markers regulated by *Shh* at the vertebrate ZLI are also expressed in similar topological arrangements at the *S. kowalevskii* proboscis–collar boundary. In vertebrates, *Pax6* and *Dlx2* are expressed anterior to the ZLI in the prethalamus (Fig. 2p, s), where their expression requires *Shh*^{17,18}. Similarly, *S. kowalevskii pax6* and *dlx* are expressed anterior to *hh* at the proboscis base (Fig. 2n, o, q, r, v). In vertebrates, *Foxa2* is expressed in the ZLI itself¹⁷, and in *S. kowalevskii*, *foxa* is expressed at the proboscis–collar boundary (Fig. 2t, u). Targeting *hh* function in *S. kowalevskii* by injecting *hh* siRNA downregulated *dlx* at the proboscis–collar boundary (Fig. 2y, z). However, these embryos had other strong defects in anteroposterior and dorsoventral patterning, probably owing to *hh* having additional roles that complicated the assessment of *hh* function at the proboscis–collar boundary (Supplementary Fig. 3). To reduce early pleiotropic effects, we treated embryos with the Hh signalling inhibitor cyclopamine⁴⁰ from the end of gastrulation through double-groove stage. Treating embryos with 50 μ M cyclopamine downregulated *dlx* at the proboscis–collar boundary with limited effects on midline expression and general morphology (Fig. 2a', b'), suggesting that Hh signalling from the proboscis–collar boundary regulates *dlx*. The prominent similarities in expression of ZLI marker homologues in *S. kowalevskii* and vertebrates, and the conserved functions for *otx* and *hh*, suggest that an ancestral signalling centre homologous to the ZLI was present in early deuterostomes.

An IsO-like regulatory program in hemichordates

The IsO is located at the midbrain–hindbrain boundary (MHB) and is defined molecularly by abutting domains of FGF8 and WNT1, which induce and pattern adjacent neural structures^{13,13,14,42}. The search for a pre-vertebrate IsO has focused on expression patterns of these ligands along with orthologues of the transcription factors *otx*, *gbx*, *en* and *pax2/5/8*, whose combinatorial expression patterns in vertebrates define a molecular territory unique to the MHB (Fig. 3b, d, g, i, m, p–r). In amphioxus, CNS expression of *fgf8/17/18* is restricted to the

anterior cerebral vesicle^{19,24}, and *wnt1* is not expressed in the CNS²⁵. In *Ciona intestinalis*, *fgf8/17/18* is expressed in the larval visceral ganglion where it regulates *en* and *pax2/5/8* to delineate the sensory vesicle and neck regions²⁰, suggesting that at least a partial IsO-like signalling centre pre-dates vertebrates. However, *wnt1* is absent in the *C. intestinalis* genome, making it difficult to infer the full extent of this ancestral centre. In *Drosophila melanogaster*, *otx* and *gbx* orthologues are expressed in patterns similar to those at the MHB, but the absence of compelling similarities in the expression of *fgf8/17/18*-related genes, and repeated expression of *wnt1* and *en* at parasegmental boundaries¹² weakens the idea of homology with the vertebrate IsO. To assess the presence of an IsO-like region in hemichordates, we investigated expression patterns for *S. kowalevskii* homologues of vertebrate MHB markers (Fig. 3a–t). We found that at double-groove stage (36 h), *fgf8/17/18* and *wnt1* are expressed in adjacent ectodermal bands in the anterior trunk with *wnt1* expressed anterior to *fgf8/17/18* (Fig. 3a, c, e); a topology similar to the vertebrate IsO (Fig. 3b, d). In vertebrates, abutting domains of *otx* and *gbx* genes position the IsO (ref. 3), and opposing *otx* and *gbx* domains are also found in protostomes¹² and amphioxus^{8,9}, indicating that this pattern is ancestral to bilaterians. A reassessment of *otx* and *gbx* expression patterns in *S. kowalevskii*⁷ revealed that they are also expressed in adjacent domains at the collar–trunk coelom boundary, with *gbx* expressed in the ectoderm between the most posterior *otx* domains (Fig. 3f, h, j). However, the spatial arrangements of *otx* and *wnt1*, and *gbx* and *fgf8/17/18*, are reversed in *S. kowalevskii* and vertebrates (Fig. 3t) indicating that divergent mechanisms position homologous IsO-like regions in these groups. In vertebrates, *pax2*, *pax5* and *pax8*, and *en1* and *en2* are co-expressed at the MHB (Fig. 3m, p–r). However, expression of *pax2/5/8* and *en* does not overlap in *S. kowalevskii* (Fig. 3n, o) or invertebrate chordates^{9,20}, which suggests that regulation of *en* genes by *pax2/5/8* genes is a novel feature of vertebrate development. Beyond similarities in expression of signalling molecules and transcription factors, we found that the catecholaminergic neuron marker tyrosine hydroxylase is co-expressed

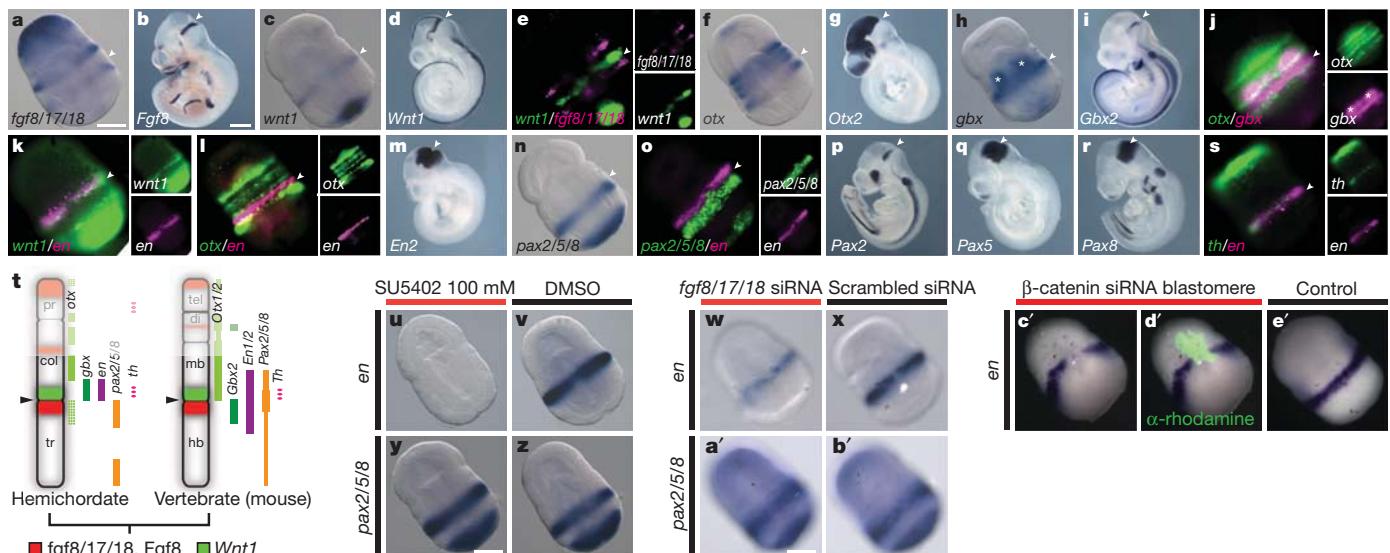


Figure 3 | An IsO-like signalling centre in *S. kowalevskii*. **a–s.** *In situ* hybridizations for *S. kowalevskii* and mouse homologues of MHB markers. Arrowheads mark the *S. kowalevskii* collar–trunk coelom boundary and the mouse IsO. **a.** *S. kowalevskii* *fgf8/17/18* expression. **b.** Mouse *Fgf8* expression. **c.** *S. kowalevskii* *wnt1* expression. **d.** Mouse *Wnt1* expression. **e.** Double FISH showing *S. kowalevskii* *wnt1* (green) expressed directly anterior to *fgf8/17/18* (magenta). **f.** *S. kowalevskii* *otx* expression. **g.** Mouse *Otx2* expression. **h.** *S. kowalevskii* *gbx* expression. Asterisks denote endodermal domains. **i.** Mouse *Gbx2* expression. **j.** Double FISH for *S. kowalevskii* *otx* (green) and *gbx* (magenta). **k.** Double FISH for *S. kowalevskii* *wnt1* (green) and *en* (magenta). **l.** Double FISH for *S. kowalevskii* *en* and *otx*. **m.** Mouse *En2* expression. **n.** *S. kowalevskii* *pax2/5/8* expression. **o.** Double FISH showing *S. kowalevskii* *pax2/5/8* and *en* expression. **p–r.** Expression of mouse *Pax2* (**p**), *Pax5* (**q**) and *Pax8* (**r**). **t.** Summary of anteroposterior expression topologies in hemichordates and mice. Anterior to top. **u–z.** *en* expression is reduced in an embryo that has been treated with SU5402. **v.** *en* expression in a DMSO-treated control embryo. **w.** *en* expression in an embryo injected with *fgf8/17/18* siRNA. **x.** *en* expression in an embryo treated with SU5402. **y.** *pax2/5/8* expression in a DMSO-treated control embryo. **z.** *pax2/5/8* expression in an embryo injected with *fgf8/17/18* siRNA. **a'–b'.** *pax2/5/8* expression in an embryo injected with a control siRNA. **c'–e'.** *en* is not expressed in descendants of a blastomere injected with β -catenin siRNA. **d'.** Merged darkfield and fluorescence images showing the location of the β -catenin-deficient clone (green). **e'.** Wild-type *en* expression. Scale bars, 100 μ m in *S. kowalevskii* embryos; 1 mm in mice.

with *en* in the *S. kowalevskii* posterior collar (Fig. 3s) similar to vertebrates⁴¹, raising the possibility that this neuronal population is homologous to vertebrate midbrain dopaminergic neurons.

Functional assays further support the deep deuterostome ancestry of a MHB-like genetic module. In vertebrates, *fgf8* mediates the organizing abilities of the IsO and maintains expression of other MHB markers^{1,3,13,14}. To assess similar requirements for FGFs in hemichordates, we first treated embryos with 100 μ M SU5402 at the end of gastrulation to suppress FGF signalling without perturbing earlier patterning events. Expression of *en* was strongly reduced in SU5402-treated embryos (Fig. 3u, v), whereas *pax2/5/8* was unaffected (Fig. 3y, z), compared to control embryos treated with dimethylsulphoxide (DMSO). To test specifically for a role of *fgf8/17/18* in regulating collar–trunk patterning, we injected fertilized oocytes with *fgf8/17/18* siRNA. Knockdown of *fgf8/17/18* reduced *en* expression (Fig. 3w, x) but had no effect on *pax2/5/8* expression (Fig. 3a', b') relative to embryos injected with a scrambled control siRNA. The absence of any effect on *pax2/5/8* expression in these experiments highlights differences in gene regulation downstream of *fgf8/17/18* homologues between hemichordates and chordates^{3,20}.

In vertebrates, *Wnt1* is required to maintain expression of *en* genes at the MHB^{3,42}. To assess local Wnt functions at the *S. kowalevskii* collar–trunk boundary, we injected single blastomeres with β -catenin siRNA at early cleavage stages to suppress Wnt signalling in small patches of collar–trunk ectoderm. Clonal descendants of injected blastomeres failed to express *en* (Fig. 3c'–e'), suggesting a similar role for Wnts in regulating *en* genes at the MHB and collar–trunk boundary.

Discussion

Extensive similarities in expression patterns of signalling-centre markers and conserved functions for FGF, Wnt and Hh signals in *S. kowalevskii* and vertebrates provide compelling evidence that signalling centres

homologous to the ANR, ZLI and IsO were parts of an ancient genetic regulatory scaffold that pre-date the morphological innovations of vertebrates (Fig. 4). Therefore, assembly of these genetic networks did not trigger morphological novelties of the brain. Instead, early vertebrate brain evolution involved modifying and elaborating ancestral signalling centres to pattern novel structures within a highly conserved gene regulatory framework for anteroposterior ectodermal patterning⁷.

Widespread losses of signalling-centre components in invertebrate chordates and the unresolved nature of the ancestral deuterostome nervous system present a challenge for inferring when the ANR, ZLI and IsO genetic circuits were first deployed to regulate CNS patterning specifically, and to what extent these integrations could have been associated with origins of vertebrate novelties. In *S. kowalevskii*, signalling-centre programs are deployed at stages when the nervous system is still circumferentially organized⁷ and in body regions that have been described as containing components of the adult peripheral, rather than central, nervous system in the hemichordate *Ptychoderida flava*⁴³. Notably, similar to their roles in the vertebrate brain, the ZLI and IsO-like signalling centres are located at morphological boundaries in *S. kowalevskii*, suggesting ancient roles in demarcating ectodermal divisions. We propose that rather than having evolved for CNS patterning, the most ancient role for signalling centres was in general body plan regionalization.

This work also highlights that basal chordates have not retained all ancestral chordate characters and have undergone substantial independent evolutionary changes. This is generally accepted for urochordates, but based on available data, cephalochordates are often considered to be the most informative extant group for reconstructing ancestral chordate characters^{8,9,24}. Our data provide clear evidence that secondary losses of complex developmental mechanisms have occurred in cephalochordates, and that overlooking this possibility

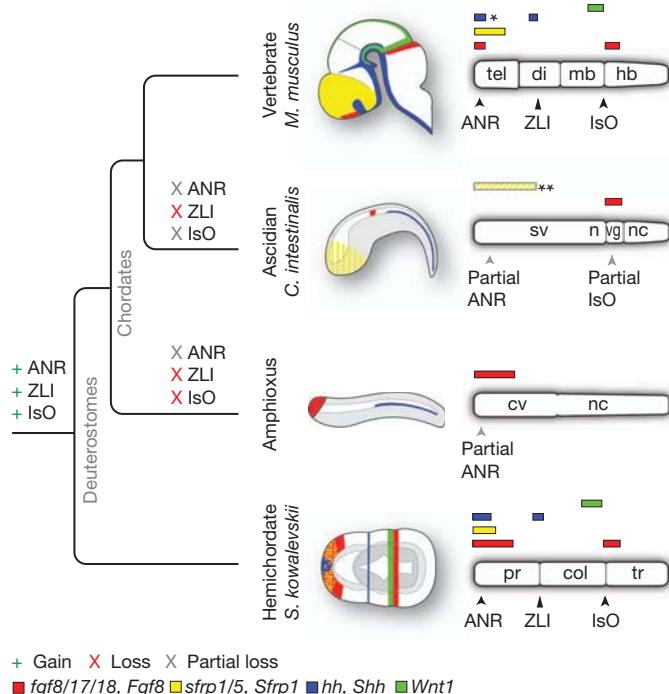


Figure 4 | Evolutionary gain and loss of ANR, ZLI and IsO-like genetic programs. Schematic diagrams depicting the expression of *Fgf8*, *Sfrp1*, *Shh* and *Wnt1* homologues in the mouse brain and ectoderm of *C. intestinalis*, amphioxus and *S. kowalevskii*. Embryos are oriented with their anterior side to the left of the image and their dorsal side to top. Bar diagrams are oriented with the anterior side to the left of the image. Diagrams depict only expression domains that are related to signalling components of vertebrate CNS signalling centres. cv, cerebral vesicle; n, neck; nc, nerve cord; sv, sensory vesicle; vg, visceral ganglion. Diagrams are not to scale. Single asterisk indicates that *Shh* is expressed in the medial ganglionic eminence, near the ANR. Double asterisk indicates that *sfrp1/5* is expressed in the *C. intestinalis* anterior ectoderm from the 64-cell stage up to neurulation but is then downregulated in the anterior ectoderm and CNS (shown as yellow stripes).

can inflate the numbers of putative vertebrate novelties. Although divergence of ANR, ZLI and IsO-like genetic programs in extant invertebrate chordate lineages may have been associated with the loss or modification of anterior ectodermal structures of the CNS and head, this is difficult to test. On the basis of these new observations, re-analysis of early chordate fossils that accommodates scenarios of greater anterior complexity in stem chordates may be informative. However, in the absence of additional fossil data, predictions of the morphological consequences of developmental genetic losses based on molecular data alone are unreliable.

This study provides a compelling example of the challenges associated with identifying key developmental genetic innovations responsible for morphological innovations at macroevolutionary scales. The unexpected presence of ANR, ZLI and IsO-like programs in *S. kowalevskii* highlights that basing outgroup choice solely on morphological criteria can lead to erroneous conclusions about links between morphological and developmental genetic characters: although by almost all morphological criteria amphioxus shares more similarities with vertebrates than do hemichordates, our data support the hypothesis that in certain cases hemichordates will be a more informative group than basal chordates for reconstructing stem chordate characters and understanding the origins of vertebrate developmental genetic processes. Additional data from protostomes, especially lophotrochozoans, will be required to assess whether the ANR, ZLI and IsO genetic programs are unique deuterostome features or have even deeper bilaterian origins.

Our findings highlight the importance of broad phylogenetic sampling, including morphologically divergent outgroups, to identify

gene regulatory innovations responsible for evolutionary changes in body plans. With growing use of novel model organisms, it seems that secondary losses of complex developmental regulatory characters may occur commonly^{44–46}, with consequences for morphological evolution that are still poorly understood. Conversely, the presence of complex developmental modules that regulate morphologically disparate structures in distantly related lineages suggests a loose coupling between morphological and gene regulatory evolution over macroevolutionary timescales, and highlights the difficulties of reconstructing ancestral morphological characters from molecular genetic data.

METHODS SUMMARY

Gravid *S. kowalevskii* were collected at Waquoit Bay National Estuarine Research Reserve near Woods Hole, Massachusetts, and maintained at the Marine Biological Laboratory, Woods Hole. Spawning and embryo rearing was carried out using established techniques⁷. For SU5402 and cyclopamine treatments, embryos were raised either in inhibitor, or in an equivalent concentration of DMSO, diluted in 0.2 µm-filtered sea water. siRNA microinjections were performed as described previously^{27,29} using calcein or lysinated 10,000 molecular weight (MW) tetramethylrhodamine dextran tracers. Descendants of injected blastomeres were detected by anti-tetramethylrhodamine immunofluorescence. siRNA sequences are provided in Supplementary Table 1. Procedures involving mice (*Mus musculus*) were performed under a protocol approved by the University of Chicago Institutional Animal Care and Use Committee. Embryo fixations and colorimetric *in situ* hybridizations were performed using standard protocols^{7,31}. See Methods for fluorescent *in situ* hybridization (FISH) protocol. Photographs were taken using Zeiss AxioCam MRC5 or MRM cameras on a Zeiss AxioImager.Z1 or a Discovery.V12. Images were acquired using Zeiss Axiovision 4.8 software and adjusted for colour balance and/or levels or gamma using Axiovision 4.8 or Adobe Photoshop CS3 or CS5 software. Images of experimental and control embryos were processed using the same parameters. *S. kowalevskii* homologues of vertebrate genes were identified in an expressed sequence tag (EST) library screen⁴⁷. Amino acid sequences were aligned using ClustalW⁴⁸, and putative homologues were confirmed by constructing gene trees using MrBayes 3.1.2 (refs 49,50) with parameters optimized for each gene (Supplementary Figs 4 and 5). See Supplementary Table 2 for accession numbers.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 November 2011; accepted 6 January 2012.

1. Echevarria, D., Vieira, C., Gimeno, L. & Martinez, S. Neuroepithelial secondary organizers and cell fate specification in the developing brain. *Brain Res. Brain Res. Rev.* **43**, 179–191 (2003).
2. Wilson, S. W. & Houart, C. Early steps in the development of the forebrain. *Dev. Cell* **6**, 167–181 (2004).
3. Wurst, W. & Bally-Cuif, L. Neural plate patterning: upstream and downstream of the isthmic organizer. *Nature Rev. Neurosci.* **2**, 99–108 (2001).
4. Wicht, H. & Lacalli, T. C. The nervous system of amphioxus: structure, development, and evolutionary significance. *Can. J. Zool.* **150**, 122–150 (2005).
5. Lacalli, T. C. Prospective protochordate homologs of vertebrate midbrain and MHB, with some thoughts on MHB origins. *Int. J. Biol. Sci.* **2**, 104–109 (2006).
6. Meinertzhagen, I. A., Lemaire, P. & Okamura, Y. The neurobiology of the ascidian tadpole larva: recent developments in an ancient chordate. *Annu. Rev. Neurosci.* **27**, 453–485 (2004).
7. Lowe, C. J. et al. Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**, 853–865 (2003).
8. Holland, L. Z. & Short, S. Gene duplication, co-option and recruitment during the origin of the vertebrate brain from the invertebrate chordate brain. *Brain Behav. Evol.* **72** (2008).
9. Holland, L. Z. Chordate roots of the vertebrate nervous system: expanding the molecular toolkit. *Nature Rev. Neurosci.* **10**, 736–746 (2009).
10. Irimia, M. et al. Conserved developmental expression of *Fezf* in chordates and *Drosophila* and the origin of the *Zona Limitans Intrathalamica* (ZLI) brain organizer. *EvoDevo*, **1**, 7 (2010).
11. Tomer, R., Denes, A. S., Tessmar-Raible, K. & Arendt, D. Profiling by image registration reveals common origin of annelid mushroom bodies and vertebrate pallium. *Cell* **142**, 800–809 (2010).
12. Urbach, R. A procephalic territory in *Drosophila* exhibiting similarities and dissimilarities compared to the vertebrate midbrain/hindbrain boundary region. *Neural Dev.* **2**, 23 (2007).
13. Crossley, P. H., Martinez, S. & Martin, G. R. Midbrain development induced by FGF8 in the chick embryo. *Nature* **380**, 66–68 (1996).
14. Reifers, F. et al. *Fgf8* is mutated in zebrafish cerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. *Development* **125**, 2381–2395 (1998).

15. Houart, C. *et al.* Establishment of the telencephalon during gastrulation by local antagonism of Wnt signaling. *Neuron* **35**, 255–265 (2002).
16. Paek, H., Gutin, G. & Hebert, J. M. FGF signaling is strictly required to maintain early telencephalic precursor cell survival. *Development* **136**, 2457–2465 (2009).
17. Kiecker, C. & Lumsden, A. Hedgehog signaling from the ZLI regulates diencephalic regional identity. *Nature Neurosci.* **7**, 1242–1249 (2004).
18. Scholpp, S., Wolf, O., Brand, M. & Lumsden, A. Hedgehog signalling from the zona limitans intrathalamica orchestrates patterning of the zebrafish diencephalon. *Development* **133**, 855–864 (2006).
19. Meulemans, D. & Bronner-Fraser, M. Insights from amphioxus into the evolution of vertebrate cartilage. *PLoS ONE* **2**, e787 (2007).
20. Imai, K. S., Stolfi, A., Levine, M. & Satou, Y. Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. *Development* **136**, 285–293 (2009).
21. Shimeld, S. M. The evolution of the hedgehog gene family in chordates: insights from amphioxus hedgehog. *Dev. Genes Evol.* **209**, 40–47 (1999).
22. Takatori, N., Satou, Y. & Satoh, N. Expression of hedgehog genes in *Ciona intestinalis* embryos. *Mech. Dev.* **116**, 235–238 (2002).
23. Scholpp, S. & Lumsden, A. Building a bridle chamber: development of the thalamus. *Trends Neurosci.* **33**, 373–380 (2010).
24. Bertrand, S. *et al.* Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. *Proc. Natl. Acad. Sci. USA* **108**, 9160–9165 (2011).
25. Holland, L. Z., Holland, N. N. & Schubert, M. Developmental expression of AmphibWnt1, an amphioxus gene in the Wnt1/wingless subfamily. *Dev. Genes Evol.* **210**, 522–524 (2000).
26. Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).
27. Darras, S., Gerhart, J., Terasaki, M., Kirschner, M. & Lowe, C. J. β -catenin specifies the endomesoderm and defines the posterior organizer of the hemichordate *Saccoglossus kowalevskii*. *Development* **138**, 959–970 (2011).
28. Gillis, J. A., Fritzenwanker, J. H. & Lowe, C. J. A stem-deuterostome origin of the vertebrate pharyngeal transcriptional network. *Proc. R. Soc. B* **279**, 237–246 (2012).
29. Lowe, C. J. *et al.* Dorsoventral patterning in hemichordates: insights into early chordate evolution. *PLoS Biol.* **4**, e291 (2006).
30. Shimamura, K. & Rubenstein, J. L. Inductive interactions direct early regionalization of the mouse forebrain. *Development* **124**, 2709–2718 (1997).
31. Fukuchi-Shimogori, T. & Grove, E. A. Neocortex patterning by the secreted signaling molecule FGF8. *Science* **294**, 1071–1074 (2001).
32. Walshe, J. & Mason, I. Unique and combinatorial functions of Fgf3 and Fgf8 during zebrafish forebrain development. *Development* **130**, 4337–4349 (2003).
33. Garel, S., Huffman, K. J. & Rubenstein, J. L. Molecular regionalization of the neocortex is disrupted in *Fgf8* hypomorphic mutants. *Development* **130**, 1903–1914 (2003).
34. Mohammadi, M. *et al.* Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science* **276**, 955–960 (1997).
35. Lagutin, O. V. *et al.* Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev.* **17**, 368–379 (2003).
36. Crossley, P. H., Martinez, S., Ohkubo, Y. & Rubenstein, J. L. Coordinate expression of *Fgf8*, *Otx2*, *Bmp4*, and *Shh* in the rostral prosencephalon during development of the telencephalic and optic vesicles. *Neuroscience* **108**, 183–206 (2001).
37. Hébert, J. M. & Fishell, G. The genetics of early telencephalon patterning: some assembly required. *Nature Rev. Neurosci.* **9**, 678–685 (2008).
38. Zeltser, L. M., Larsen, C. W. & Lumsden, A. A new developmental compartment in the forebrain regulated by Lunatic fringe. *Nature Neurosci.* **4**, 683–684 (2001).
39. Scholpp, S. *et al.* *Otx1*, *Otx2* and *Irxa* establish and position the ZLI in the diencephalon. *Development* **134**, 3167–3176 (2007).
40. Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16**, 2743–2748 (2002).
41. Simon, H. H., Thuret, S. & Alberi, L. Midbrain dopaminergic neurons: control of their cell fate by the engrailed transcription factors. *Cell Tissue Res.* **318**, 53–61 (2004).
42. McMahon, A. P., Joyner, A. L., Bradley, A. & McMahon, J. A. The midbrain-hindbrain phenotype of *Wnt1*–*Wnt1*[–] mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell* **69**, 581–595 (1992).
43. Nomaksteinsky, M. *et al.* Centralization of the deuterostome nervous system predates chordates. *Curr. Biol.* **19**, 1264–1269 (2009).
44. Gavino, M. A., Reddien, P. W. & A. Bmp/Admp regulatory circuit controls maintenance and regeneration of dorsal-ventral polarity in planarians. *Curr. Biol.* **21**, 294–299 (2011).
45. Grande, C. & Patel, N. H. Nodal signalling is involved in left–right asymmetry in snails. *Nature* **457**, 1007–1011 (2009).
46. Campo-Paysaa, F., Marletaz, F., Laudet, V. & Schubert, M. Retinoic acid signaling in development: tissue-specific functions and evolutionary origins. *Genesis* **46**, 640–656 (2008).
47. Freeman, R. M. Jr *et al.* cDNA sequences for transcription factors and signaling proteins of the hemichordate *Saccoglossus kowalevskii*: efficacy of the expressed sequence tag (EST) approach for evolutionary and developmental studies of a new organism. *Biol. Bull.* **214**, 284–302 (2008).
48. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
49. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
50. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Gerhart and M. Kirschner for assistance and support, R. Freeman for bioinformatics assistance, E. Farrelly, M. Terasaki and S. Darras for technical guidance, the members of the Lowe laboratory for discussions, and G. Wray and J. Gerhart for comments on drafts of the manuscript. We also thank the staff of the Marine Biological Laboratory, the Waquoit Bay National Estuarine Research Reserve, Carl Zeiss and Nikon for assistance during our field season. This work was funded by the Searle Kinship Foundation, Brain Research Foundation and National Science Foundation grant 1049106 (C.J.L.), National Institutes of Health grant R01 HD42330 (E.A.G.) and the University of Chicago Hinds Fund (A.M.P.). A.M.P. was supported by a Marine Biological Laboratory Frank R. Lillie Fellowship, National Institute of Child Health and Development institutional training grant 1T32HD055164-01A1, and National Institute of Neurological Disorders and Stroke pre-doctoral fellowship 1F31NS074738-01A1. J.A. was supported by a National Science and Engineering Research Council of Canada pre-doctoral grant.

Author Contributions A.M.P., C.J.L. and J.A. conceived the project. A.M.P. and C.J.L. performed the hemichordate experiments and wrote the paper. E.E.M. and S.A. performed mouse experiments, and E.A.G. edited the paper. All authors discussed and commented on the data.

Author Information *S. kowalevskii* gene sequences have been deposited in GenBank, and accession numbers are provided in Supplementary Table 2. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.J.L. (clowe@stanford.edu).

METHODS

Embryo procurement. Gravid *S. kowalevskii* adults were collected from the intertidal zone of sandy estuaries near Woods Hole, Massachusetts, and maintained in flow-through seawater tables at the Woods Hole Marine Biological Laboratory. Spawning and embryo rearing were carried out using established techniques⁷. Mouse embryos were obtained using standard protocols approved by the University of Chicago Institutional Animal Care and Use Committee.

Experimental manipulations. SU5402 (Calbiochem) and cyclopamine (Calbiochem) were resuspended in minimal volumes of DMSO and diluted in 0.2 μ m-filtered sea water for treatments. Embryos were raised at room temperature. Control embryos were obtained from the same batches and raised in an equivalent concentration of DMSO. siRNA microinjections were performed as described previously^{27,29}. siRNAs were resuspended at 100 mM and injected at 1/10 concentration in a suspension buffer²⁷ including either calcein or lysinated 10,000 MW tetramethylrhodamine dextran (Molecular Probes). Unsuccessfully injected embryos, or embryos with abnormal cleavage patterns, were discarded. siRNAs were ordered from Ambion or Qiagen, and sequences are provided in Supplementary Table 1.

In situ hybridizations. Colorimetric *in situ* hybridizations were performed as described previously^{7,31}. Clonal descendants of injected blastomeres were identified after *in situ* hybridization using anti-tetramethylrhodamine (Molecular Probes) immunofluorescence. For FISH, pre-hybridization steps were performed using our standard protocol. For double FISH, embryos were simultaneously hybridized with antisense RNA probes for both genes labelled separately with digoxigenin-11-UTP (Roche) or fluorescein-12-UTP (Roche). Owing to a reduced sensitivity of FISH, we typically used two to three times more probe compared to colorimetric methods. We used our standard protocol for post-hybridization washes and blocking steps. Embryos were then incubated in anti-digoxigenin-POD or anti-fluorescein-POD antibodies (Roche) diluted 1:200 in blocking solution (2% Roche blocking reagent, 1 \times MAB) for 4 h at room temperature on a rotary shaker. Embryos were washed four times in 1 \times MAB for 30 min and once for 1 h at room temperature, or overnight at 4 °C. Embryos were then washed in 0.1 M imidazole (Sigma) in 1 \times PBS for 10 min at room temperature and probes were detected using a Tyramide Signal Amplification (TSA) Plus

kit (Perkin-Elmer). Embryos were rinsed in 1 \times amplification diluent and incubated in cyanine-3 or cyanine-5 tyramide diluted 1:50 in 1 \times amplification diluent for 45 min on a rotary shaker at room temperature. Embryos were washed three times in detergent solution (1% triton X-100, 1% SDS, 0.5% sodium deoxycholate, 50 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 150 mM sodium chloride) for 20 min at 60 °C. When detecting a second gene, we washed the embryos once for 20 min in solution X (50% formamide, 2 \times SSC and 1% SDS) at 60 °C. Embryos were then washed three times in 1 \times MAB for 5 min and returned to the blocking step.

For genes with low signal under standard FISH methods, we performed an additional 2,4-dinitrophenol (DNP) amplification step. After primary antibody incubation and 1 \times MAB washes, embryos were incubated in DNP amplification reagent that was diluted 1:50 in 1 \times amplification diluent (Perkin-Elmer) for 5–10 min at room temperature on a rotary shaker. Embryos were washed four times in 1 \times MAB for 30 min and once for 1 h at room temperature, re-blocked for 1 h and incubated in anti-DNP-HRP antibody (Perkin-Elmer) that was diluted 1:200 in blocking solution for 4 h at room temperature on a rotary shaker. Embryos were washed four times in 1 \times MAB for 30 min and once for 1 h at room temperature. Expression was then detected using a TSA Plus kit (Perkin-Elmer) as described above.

Photography. Embryos were photographed using Zeiss AxioCam MRc5 or MRm cameras on Zeiss AxioImager.Z1 or Discovery.V12 microscopes. For optical clearing, *S. kowalevskii* embryos were dehydrated into methanol, cleared using Murray's clearing reagent (1:2 benzyl alcohol to benzyl benzoate) and mounted in permount (Fisher Scientific). Cleared embryos were imaged under differential interference contrast optics and uncleared embryos were photographed in 1 \times PBS or 1 \times MAB on agarose-coated dishes. Images were acquired using Zeiss Axiovision 4.8 software and adjusted for colour balance and/or levels or gamma using Zeiss Axiovision 4.8 or Adobe Photoshop CS3 or CS5 software. Images of experimental and control embryos were processed using the same parameters.

Gene identification. *S. kowalevskii* homologues of vertebrate genes were identified in an EST library screen⁴⁷, amino acid sequences were aligned using ClustalW⁴⁸ and putative homologues were confirmed by constructing gene trees using MrBayes 3.1.2 (refs 49,50) with parameters optimized for each gene (Supplementary Figs 4 and 5). See Supplementary Table 2 for accession numbers.

Genome-wide structure and organization of eukaryotic pre-initiation complexes

Ho Sung Rhee¹ & B. Franklin Pugh¹

Transcription and regulation of genes originate from transcription pre-initiation complexes (PICs). Their structural and positional organization across eukaryotic genomes is unknown. Here we applied lambda exonuclease to chromatin immunoprecipitates (termed ChIP-exo) to examine the precise location of 6,045 PICs in *Saccharomyces*. PICs, including RNA polymerase II and protein complexes TFIIA, TFIIB, TFIID (or TBP), TFIIE, TFIIF, TFIH and TFIK were positioned within promoters and excluded from coding regions. Exonuclease patterns were in agreement with crystallographic models of the PIC, and were sufficiently precise to identify TATA-like elements at so-called TATA-less promoters. These PICs and their transcription start sites were positionally constrained at TFIID-engaged downstream +1 nucleosomes. At TATA-box-containing promoters, which are depleted of TFIID, a +1 nucleosome was positioned to be in competition with the PIC, which may allow greater latitude in start-site selection. Our genomic localization of messenger RNA and non-coding RNA PICs reveals that two PICs, in inverted orientation, may occupy the flanking borders of nucleosome-free regions. Their unambiguous detection may help distinguish bona fide genes from transcriptional noise.

Assembly of the PIC and its post-assembly control are critical early steps in the transcription of eukaryotic genes. TBP (TATA-binding protein) arrives at most promoters as part of the multi-subunit TFIID complex that includes TBP-associated factors (TAFs)¹. Together these proteins help recruit RNA polymerase (Pol) II and its entourage of general transcription factors (GTFs) to the transcription start sites (TSSs) of genes^{2–4}. These PICs assemble in nucleosome-free promoter regions (NFRs) that are flanked by an upstream –1 nucleosome and a downstream +1 nucleosome⁵. PICs have largely been defined biochemically using purified GTFs at a few model genes^{2–4}, but little is known about their assembly and organization *in vivo*, particularly at near-base-pair resolution on a genome-wide scale.

An oddity of TBP is that when it is part of the TFIID complex, it tends to bind promoters that lack the TATA box consensus TATAWAWR (W indicates A/T; R indicates A/G)⁶. Approximately 80–90% of all *Saccharomyces* genes are thus designated as ‘TATA-less’, and have a predominant PIC assembly mechanism and chromatin architecture that differs substantially from those in the ‘TATA box’ class of genes^{6–9}. So far, no TBP-binding motif has been identified at TATA-less promoters, and so the origins of TFIID-promoter specificity have been rather enigmatic¹⁰. When TBP is not part of the TFIID complex, the SAGA complex directs TBP to TATA-box-containing Pol II promoters^{11–13}.

TFIIA and TFIIB clamp TBP to DNA, and make DNA contacts immediately upstream and downstream of the TATA box. TFIIB is a linchpin between TBP and Pol II^{14,15}. Its intimate contact with Pol II directs how far downstream Pol II productively initiates transcription^{16,17}. TFIIF enhances the interaction of Pol II with TFIIB, assists in recruiting TFIIE, and promotes downstream elongation events^{3,18}. TFIIE then stimulates DNA strand separation by Pol II at the TSS, and enhances the activity of TFIH. TFIH holoenzyme is multi-functional, having ATP-dependent helicase (Ssl2 and Rad3) and kinase (Kin28) activities that reside on biochemically separable sub-complexes (TFIIF and TFIK, respectively), both of which are key to efficient open complex formation and transcription initiation^{19–21}.

We examined the structural organization of PICs and their specificity on a genomic scale using ChIP-exo²². This novel strategy substantially improved mapping resolution and eliminated many false positives. The exonuclease processively degrades a DNA strand in the 5'-3' direction until a crosslinking point is encountered (Supplementary Fig. 1a). The crosslinking inefficiency inherent to ChIP allows multiple crosslinking points to be detected in a population by deep sequencing. When applied to the GTFs on a genomic scale, we obtained detailed and comprehensive information on PIC structure and genomic organization.

Genome-wide PIC structure

We applied ChIP-exo genome-wide to Pol II and each GTF (Fig. 1a), and verified binding for TFIIB by locus-specific PCR using a series of tiled primers (Supplementary Fig. 1b). When exonuclease stop sites were mapped over all annotated mRNA promoters that contained a TATA box consensus, a distinctive pattern was observed (Fig. 1b). Importantly, each GTF displayed a strand-specific composite pattern of exonuclease stop sites that occurred only when TATA boxes, but not TSSs (Supplementary Fig. 2a and data not shown), were aligned, indicating that PICs are positioned with respect to the TATA box.

For TFIIB, we detected four DNA crosslinking points (pairs of exonuclease stops), designated B1–B4 (Fig. 1c). The TATA box was precisely centred between B1 and B2, which were separated by 20 ± 3 base pairs (bp). Crosslinking point B3 and the diffuse B4 region indicate that TFIIB further crosslinked over a broad region downstream of the TATA box towards the TSS. We compared the four TFIIB crosslinking points to crystallographic-based models of open and closed TBP-TFIIB-Pol-II-promoter complexes (Fig. 1c)^{14,15}. An important caveat of the crystallographic models is they were built from multiple independent structures of truncated TFIIB-TBP-TATA, TFIIB-Pol II and Pol II elongation complexes. Thus, the combined structures represent a hypothetical organization.

Within the modelled closed and open structures, crosslinking site B1 precisely (± 3 bp) mapped to where the TFIIB carboxy-terminal core straddles the upstream DNA-binding stirrup of TBP (11 bp

¹Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

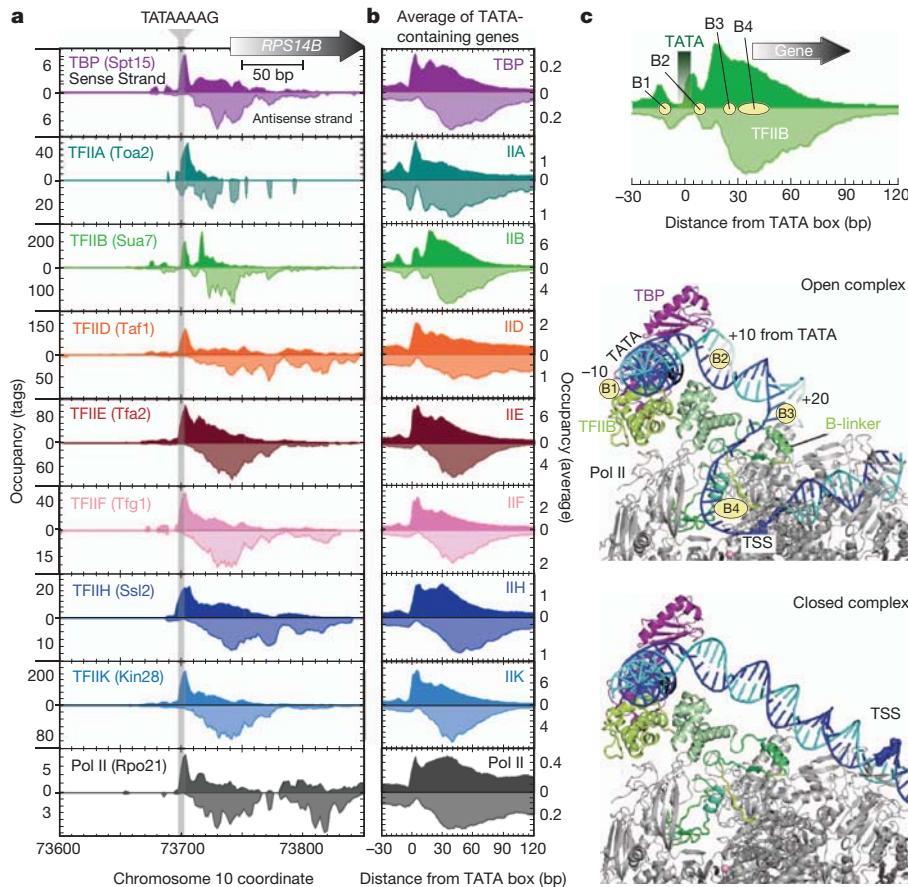
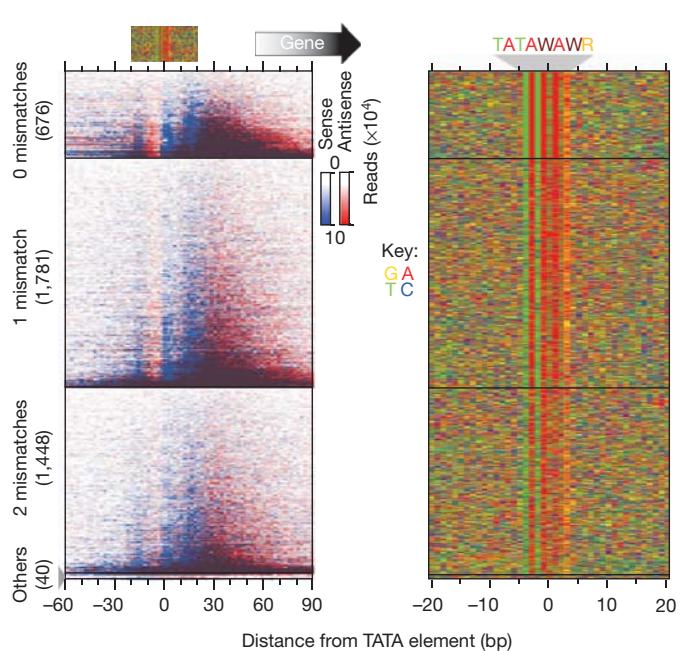


Figure 1 | Genome-wide structural organization of PICs. **a**, Raw ChIP-exo tag distribution for GTFs and Pol II around the *RPS14B* gene. Filled plots represent unfiltered 5' ends of sequencing tags on the sense (darker fill) and antisense strand (lighter fill). **b**, Average GTF and Pol II occupancy around the TATA box of 676 annotated mRNA genes. Plots are as in panel **a**.

upstream of the TATA box midpoint). B2 mapped precisely (± 3 bp) to where the TFIIB core amino-terminal cyclin fold encounters DNA just downstream of TBP's other stirrup (9 bp downstream of the TATA box midpoint). B3 mapped to where the TFIIB linker region is closest to the DNA, which was 19 bp downstream of the TATA box midpoint. B4 corresponded to a broad region defined by close proximity of the TFIIB reader (or finger) domain to single-stranded DNA within the modelled open complex, but was not evident within the closed complex. Similar broad regions of crosslinking were observed with the other GTFs (Fig. 1b), and may reflect indirect crosslinking. Support that these PICs represent open complexes is provided by permanganate reactivity studies of the *GAL1*, *GAL10* and *HSP82* loci^{23,24}. Taken together, the entirety of the genomic crosslinking sites observed with the GTFs and Pol II fits remarkably well with the crystallographic models of the PIC open complex^{14,15}, and with many aspects of *in vitro* chemical crosslinking of these proteins^{19,25–27}.

TATA-like elements at TATA-less genes

An apparent paradox of so-called TATA-less promoters is their utilization of TFIID, which has long been described as the TATA-box-binding complex⁴. However, it is unclear whether the TBP subunit of TFIID recognizes specific DNA sequences at TATA-less promoters. Inasmuch as TBP is expected to be found at all TATA-less promoters, and motif searching algorithms failed to identify candidate TBP-binding sites, we instead opted to search for sequence elements with up to two mismatches to the TATAWAWR consensus. We also limited our search to measured PIC locations. Remarkably, 99% of the PICs at TATA-less promoters contained a sequence having two or less mismatches to the TATA box consensus (Fig. 2). We refer to these mismatched elements as



'TATA-like', as they did not form a consensus, whereas those conforming to the consensus retain the 'TATA box' designation. We refer to the two elements together as 'TATA elements'.

To assess whether TFIIB was positioned around these TATA-like elements in a canonical manner as seen at bona fide TATA boxes, strand-specific ChIP-exo tags were plotted around each element, separated into panels by 0, 1 or 2 mismatches to the TATA box consensus (Fig. 2). Notably, regardless of its occupancy level, the distribution of TFIIB crosslinking and thus its positioning relative to these TATA-like elements was quite similar to the positioning observed at bona-fide TATA boxes. When the other GTFs were examined, their patterns relative to TATA-like elements were also similar to those found at TATA boxes (Supplementary Fig. 2b), although some downstream differences were observed (addressed later). Thus, as previously seen at the three yeast TATA-box-containing genes *GAL1*, *GAL10* and *HSP82* (refs 23, 24), and in mammalian *in vitro* systems²⁸, at least the upstream portion of most PICs are positioned with respect to resident TATA elements nearly identically, regardless of their Pol II promoter classification as TATA-box containing or TATA-less. Although the 'TATA-less' designation may be a misnomer, this class of genes is not simply a slight variation of the TATA class, but instead has predominant regulation by TFIID versus SAGA, positive versus negative regulation by chromatin, and lower plasticity of expression^{6–9}.

TATA-less TSS positioning by nucleosomes

Permanganate reactivity experiments detect open complex formation upstream of the TSS at the *GAL1*, *GAL10* and *HSP82* TATA-box-containing promoters^{23,24}. These and other studies^{17,29} have led to the notion that Pol II scans downstream from the open complex to find the TSS. In agreement with this being a general mechanism, we find that PICs of TATA-box-containing genes generally reside upstream of the TSS (Supplementary Fig. 2a).

TAfs are largely depleted at TATA-box-containing promoters, although they are not entirely absent (Fig. 3a). The low level of Taf1 that was present tended to be positioned similarly to TBP and other GTFs (Fig. 1b and Supplementary Fig. 2a, b). In contrast, Taf1-enriched/TATA-less promoters (which are related, as shown in Supplementary Fig. 3) showed additional interactions downstream of the TSS that exactly coincided with the size and location of the +1 nucleosome (Fig. 3a). Indeed, Taf1 displayed a more uniform positioning pattern in relation to the TSS and +1 nucleosome than to TATA elements, which suggests that at least part of the TFIID

TAF complex engages and is positioned by the +1 nucleosome at TATA-less promoters. Consistent with this, Bdf1, which is considered to be a missing piece of Taf1 (ref. 30), binds to the +1 nucleosome³¹. Furthermore, Bdf1 showed a nearly identical ChIP-exo pattern to that of Taf1 (Supplementary Fig. 4). TFIID–nucleosomal interactions have also been reported in mammalian systems, although the details may differ³².

If TFIID binds simultaneously to both the +1 nucleosome and a TATA element, then the intervening Pol II would seem to be fenced in by TFIID, thereby limiting its ability to scan DNA. This model predicts that the TSS would reside closer to the TATA element and be positionally restricted relative to the +1 nucleosome, compared to Taf1-depleted/TATA-box-containing promoters. Indeed, the TSS at TATA-less promoters resided ~10–20 bp closer to the TATA element than at TATA-box-containing promoters (Supplementary Fig. 2a).

We also compared the position of TATA elements and TSSs in relation to the +1 nucleosome. We separately examined individual Taf1-depleted/TATA-box-containing and Taf1-enriched/TATA-less promoters (Fig. 3b). Strikingly, at the Taf1-enriched promoters, the TSS was tightly positioned at the border between the 5' NFR and the +1 nucleosome in comparison to Taf1-depleted/TATA-box-containing promoters. The latter had TSSs distributed across the adjacent nucleosome position, and these nucleosomes were relatively depleted compared to the Taf1-enriched class (Fig. 3a).

Taken together, we interpret these observations as reflecting distinct functions of the +1 nucleosome at the two classes of genes. Nucleosomes and PICs of the TFIID-enriched/TATA-less class might cooperatively assemble, in which case the nucleosome may be instructive for TSS selection by impeding Pol II scanning (Fig. 3c). In contrast, nucleosomes and PICs of the Taf1-depleted/TATA-box-containing class may competitively assemble. This may allow for the greater stochasticity or plasticity of expression that is characteristic of this class⁹, in which nucleosome loss would prime the gene for a high level of transcription. A nucleosome competition mechanism removes an impediment to Pol II scanning. Pol II could scan further, thereby allowing productive initiation at specific DNA elements³³. The transition into a scanning state may be rate limiting in the scanning cycle, as the PIC is detected upstream of the TSS.

GTF depletion in genes and their termini

Although it is clear that GTFs assemble in promoter regions of mRNA genes and disengage Pol II within ~150 bp of the TSS^{34–37}, it is less

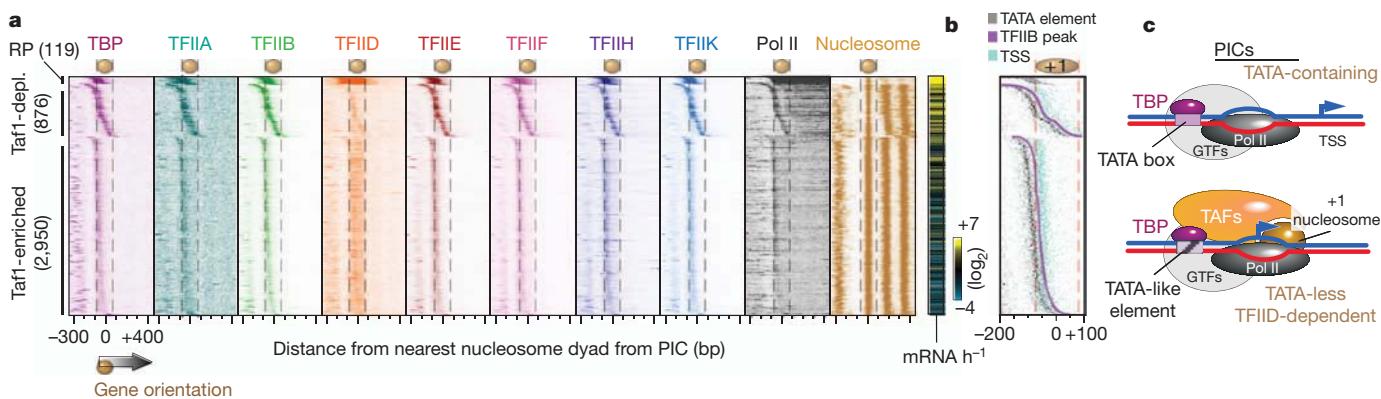


Figure 3 | PIC organization in relation to TFIID and the +1 nucleosome. **a**, GTF occupancy around the nearest nucleosome position (essentially +1) to an mRNA PIC, which were sorted by the distance between the two. Unfiltered tags on each strand were shifted in the 3' direction by a fixed distance (~8 bp depending on each GTF, 73 bp for nucleosomes), so as to reflect better the points of crosslinking. Taf1-depleted (Taf1-depl.) and Taf1-enriched genes were determined as being distinct clusters when GTF occupancies of all genes were clustered by *k*-means (see Fig. 5a). For all graphs of this type, image

resolution is less than the number of rows, resulting in some averaging and thus the appearance of less variance across adjacent rows. See Supplementary Data 2 for underlying values, which can be visualized in Treeview. RP indicates ribosomal protein genes. The right panel shows transcription frequency⁴⁸. The nucleosome borders are denoted by vertical dashed black lines. **b**, Same as panel a, but showing an overlay of TATA elements, TFIIB and TSS. **c**, Model of PIC organization at TATA-box-containing and TATA-less/TFIID-dependent genes.

clear to what extent they assemble across the body of genes, or at genes that are either transcriptionally silent or classified as 'dubious'. A whole genome view of GTFs and Pol II is presented in Fig. 4a. Remarkably, PICs were almost entirely excluded from coding regions, regardless of gene activity, whereas Pol II was enriched across gene bodies, as expected. Approximately 90% of dubious open reading frames (ORFs) lacked a canonical PIC organization or contained PICs within the ORF, and thus are unlikely to be coding. Thus, coding region PIC-driven initiation, whether in the sense or antisense direction, is infrequent on the scale of what is seen at mRNA promoters. Moreover, the observed GTF pattern makes clear that Pol II disengages all GTFs at the promoter.

Much less is known of the fate of Pol II at the ends of genes, as it undergoes termination. To examine the 3' ends of genes, without the complications associated with nearby mRNA promoters, we separated 3' ends into those having nearby 3' or 5' ends of an adjacent gene (Fig. 4b). Within terminal intergenic regions, GTFs were highly

depleted, indicating that PICs rarely exist at the 3' ends of genes at levels seen for mRNA genes (although lower levels do exist).

Remarkably, we find a highly correlated enrichment of Pol II and TFIID (Ssl2) but not TFIID (Kin28), at the end of genes within 3' NFRs (Fig. 4b). Such a physical separation of the TFIID/Ssl2 and TFIID/Kin28 submodules of holo-TFIID in genome-wide binding experiments has not previously been reported, but may be in accord with their biochemistry^{19–21}. However, Ssl2 is biochemically separable from the TFIID core, which therefore prompted us to examine additional core TFIID subunits, Ssl1 and Tfb1. Surprisingly, both were absent from the ends of genes (Supplementary Fig. 6), although they were present within PICs at promoters. These results suggest that Ssl2, a 3'-5' helicase, operates independently of holo-TFIID at the ends of genes. Consistent with a possible role of Ssl2 in transcription termination, Ssl2 has functional interactions with the Hsp90 protein chaperone³⁸, which has been implicated in the disassembly of the transcription machinery³⁹.

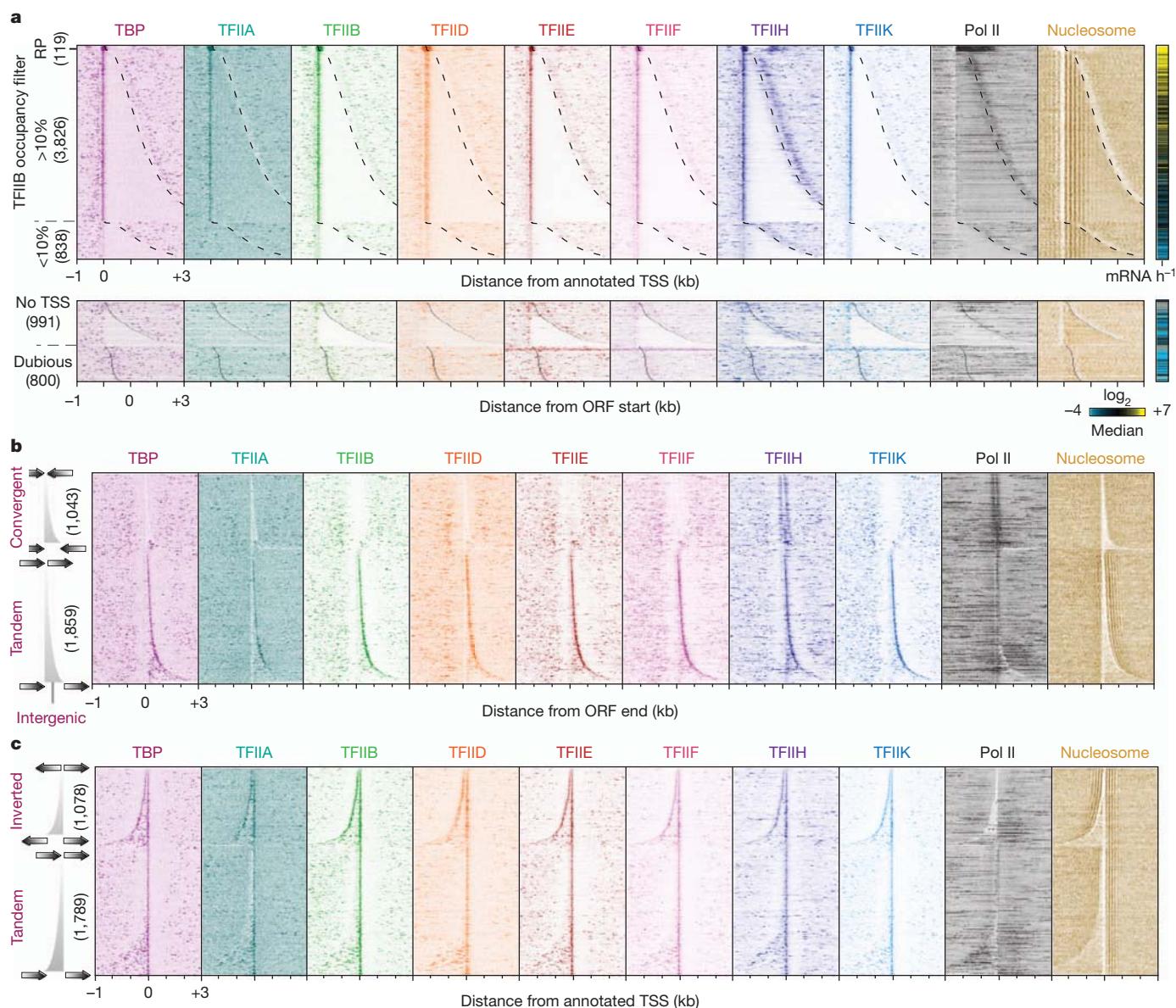


Figure 4 | Genomic view of PICs in relation to genes. **a**, GTF occupancy around transcript and ORF start sites⁴⁷, sorted by gene length. See Supplementary Data 3–5 for underlying values. Transcript or ORF ends are indicated by black dashed and solid lines, respectively. The right panel shows transcription frequency⁴⁸. **b**, GTF distribution around the 3' ends of mRNA

genes, sorted by intergenic length, and sectioned by convergent versus tandem gene-pairs. Occupancy at eight reported looped genes^{49,50} is shown in Supplementary Fig. 5. **c**, GTF distribution around the TSS of mRNA genes, sorted by intergenic length, and sectioned by inverted versus tandem gene-pairs.

Divergent transcription from distinct PICs

In contrast to coding regions, PICs were abundant in intergenic regions, far beyond what could be accounted for at mRNA promoters (Fig. 4c). Divergent transcription, in which mRNA and non-coding (nc)RNA initiation occurs within the same region but elongates in opposite directions, is well established in eukaryotes^{40,41}. However, it has been unclear whether divergent transcription originates from the same PIC site. Conceivably, the entire PIC or a portion thereof might assemble in either direction. As shown in Fig. 4c, even the shortest (~120 bp) 5' intergenic regions of mRNA genes with inverted orientation were associated with two PICs, one for each mRNA direction. Thus, divergent mRNA transcription originates from two distinct PICs, even when arising from the same NFR.

We next examined the composition and location of PICs associated with mRNA and ncRNA (variously classified as cryptic unstable transcripts (CUTs), stable uncharacterized transcripts (SUTs), Xrn1-sensitive unstable transcripts (XUTs) and antisense (AS))^{42–44}. We also examined orphan PICs, which we defined as being >160 bp from any annotated TSS or ORF start site. Nearly all had the same relative

composition of GTFs, including Taf1 depletion or enrichment (Fig. 5a), although mRNA PICs generally had higher occupancy levels. GTFs had highly correlated occupancies at all PICs (see Supplementary Fig. 7 for mRNA PICs). ncRNA PICs were generally organized around an adjacent nucleosome, as seen for mRNA PICs (Supplementary Fig. 8). Thus, all mRNA, ncRNA and orphan PICs are compositionally homogeneous with regards to the GTFs (excluding TAFs).

To visualize better the context of the low-occupancy ncRNA PICs with mRNA genes, we marked ncRNA PIC locations by their TATA element, and plotted their directionality with respect to nearby mRNA (Fig. 5b). We observed a general trend where ncRNA and mRNA PICs were positioned in opposite directions 150–200 bp apart. This places both PICs within the same NFR, and thus within the same canonical nucleosome architecture, as seen for two divergent mRNA PICs that share the same NFR (Supplementary Fig. 9).

Low-occupancy ncRNA PICs were also found towards the 3' ends of mRNA genes, of which the majority were antisense to the mRNA (Fig. 5b), and associated with low expression of the sense mRNA

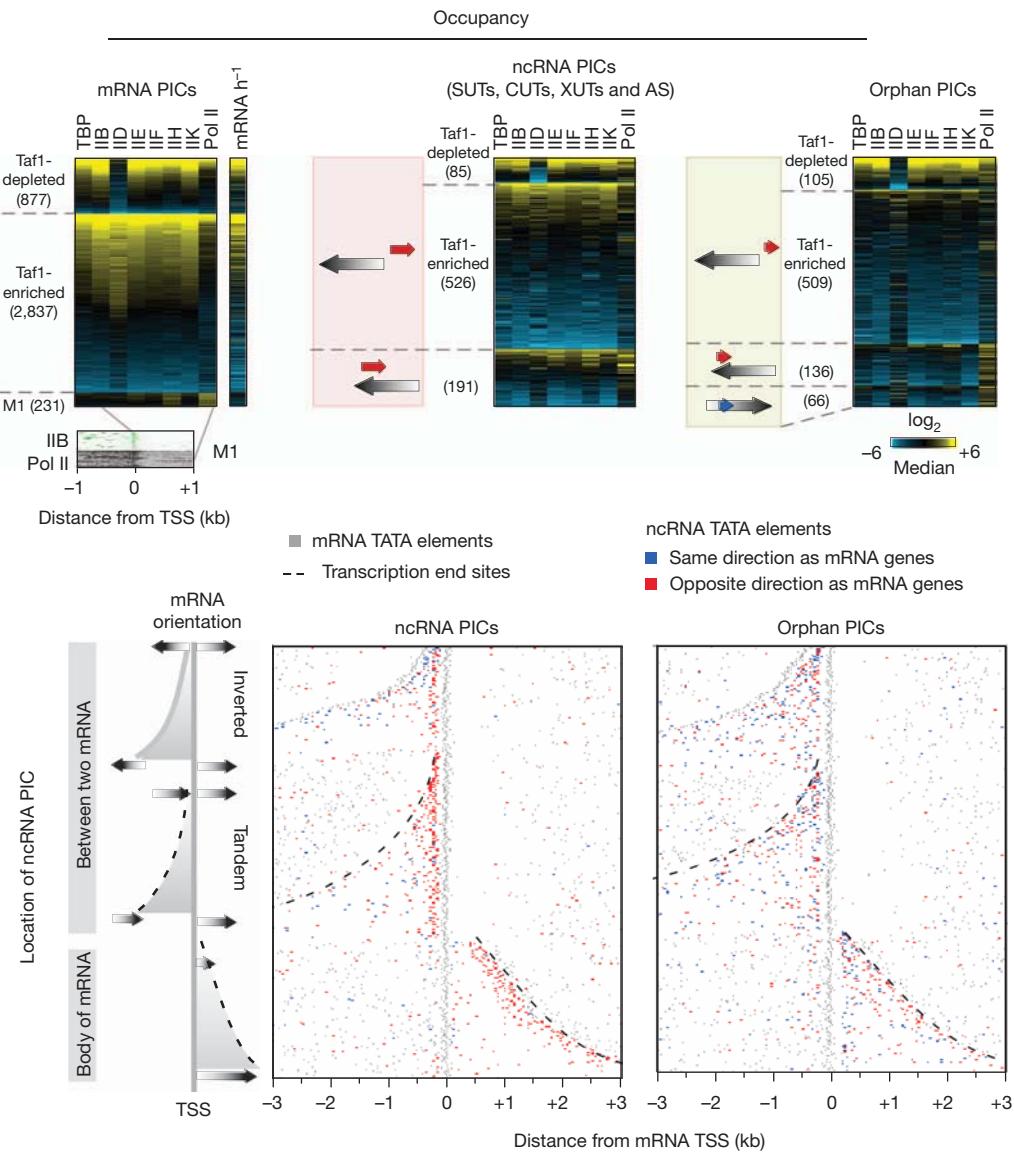


Figure 5 | Distribution of ncRNA PICs. **a**, GTF occupancy levels at PICs for mRNA, ncRNA and orphans, sectioned by *k*-means clustering. Occupancy levels of each GTF were median normalized, \log_2 transformed, then sorted by row median. The small M1 group represents a terminating polymerase originating from upstream. **b**, Distribution of ncRNA PICs relative to mRNA

genes. mRNA genes were filtered to retain only those having a nearby ncRNA-associated PIC (defined as having >10% of the genome-wide TFIIB average). Plotted are the locations of TATA elements associated with the mRNA (grey), sense-directed ncRNA (blue), and antisense-directed ncRNA (red). Additional plot details are as described in Fig. 4.

(Supplementary Fig. 10a). Thus, ncRNAs, which tend to be antisense⁴⁰, are generally associated with repression when residing in gene bodies.

In total, we identified ~6,000 PICs in rapidly growing yeast cells, in which the PICs had an occupancy level of >10% of the genome average. More than 98% of these PICs had a TATA element precisely where TBP bound. Approximately 70% of the identified PICs were associated with mRNA genes (Supplementary Fig. 10b). The remaining ~30% were divided evenly between ncRNA and orphans. At lower detection thresholds, many more low-occupancy PICs could be identified. We do not believe that they represent technical noise, as they are highly enriched in NFRs in which mRNA and ncRNA PICs are found. They might produce low levels of promoter-specific basal transcription.

Unifying principles of PICs

Our data suggest that with the exception of TAFs, PICs are compositionally homogeneous in regards to GTFs at coding and non-coding Pol II transcription units in the yeast genome. PICs differ markedly in occupancy levels, which is in accord with their transcription frequency. PICs tend to form at NFR/nucleosome borders at the 5' end (and to some extent at the 3' end) of genes, where they direct either mRNA or ncRNA transcription away from the NFR. As such, an NFR may normally accommodate two divergently oriented PICs at markedly different occupancy and transcription levels. These occupancy levels do not strictly correlate, which suggests largely independent control of the two PICs.

PIC assembly, orientation and positioning may be contributed to in part by the resident TATA element, as well as through sequence-specific factors and co-factors. TFIID-regulated promoters may rely less on TATA element strength, and more on an NFR-adjacent nucleosome for PIC assembly, orientation and positioning. The adjacent nucleosome might also serve to impede a scanning Pol II so that it can productively select a TSS at a focused position just inside the nucleosome. At SAGA-regulated promoters, which tend to contain a consensus TATA box, nucleosome occupancy may be more competitive with PIC assembly, wherein the strength of TBP/TATA interactions would be more important for PIC assembly, orientation and positioning. As such, there would be no nearby nucleosome to impede polymerase scanning, which would allow TSS selection to be controlled by other factors, including DNA sequence.

The emergent concept of ncRNA and the difficulty of distinguishing random transcriptional noise from specific initiation raise the question as to what constitutes a gene⁴⁵. The unambiguous and precise mapping of PIC locations across a genome, as described here, might help define the start of individual genes.

METHODS SUMMARY

Saccharomyces strains (BY4741) bearing TAP-tagged GTF or Pol II subunits (or untagged TBP) were grown to exponential phase in yeast extract peptone dextrose (YPD) media (30 °C to OD_{600 nm} = 0.8), then subjected to 1% formaldehyde crosslinking for 15 min at 25 °C. Cells were harvested and washed. Sonicated chromatin was prepared by standard methods. Standard ChIP methods were used, followed by lambda exonuclease treatment and library construction, as described elsewhere²². Libraries were sequenced by an ABI SOLiD sequencer. Figures displaying strand-specific sequencing tags represent the raw data without normalization to input. TFIIB peak calls were made with GeneTrack software⁴⁶. PICs ($n = 6,045$) were identified as having a TFIIB peak-pair in at least two out of four biological replicates and having ≥ 33 sequence tags (>10% of average TFIIB occupancy)²². PICs were assigned to the nearest TSS within ± 200 bp, with mRNA⁴⁷ having precedence over ncRNA. For this purpose, ORFs lacking a TSS (from the *Saccharomyces* Genome Database) were assigned a hypothetical TSS based on the genome-wide consensus. PICs of ncRNA were assigned to the nearest TSS within ± 200 bp of SUTs, CUTs, ASs and XUTs^{42–44}, with SUTs/CUTs having precedence over AS/XUTs. To assign directionality to orphan PICs, we compared nucleosome occupancy on the lower versus higher coordinate side of TFIIB locations. If the higher coordinate had higher nucleosome occupancy it was classified as 'sense', otherwise it was 'antisense'. We validated this method by

applying it to mRNA PICs, and found >91% of the assignments to be correct. We searched for TATA elements between -80 to +20 bp of the midpoint of 6,045 TFIIB-bound locations on the sense strand, first by searching for the consensus TATAAWWR, then for one and then two mismatches to the consensus. The TATA element closest to -28 bp of a TFIIB peak had precedence if multiple elements were found.

Received 8 October; accepted 20 December 2011.

Published online 18 January 2012.

1. Green, M. R. TBP-associated factors (TAFs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem. Sci.* **25**, 59–63 (2000).
2. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549–561 (1989).
3. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10**, 2657–2683 (1996).
4. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
5. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature Rev. Genet.* **10**, 161–172 (2009).
6. Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
7. Huisings, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol. Cell* **13**, 573–585 (2004).
8. Lee, T. I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* **405**, 701–704 (2000).
9. Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18**, 1084–1091 (2008).
10. Sugihara, F., Kasahara, K. & Kokubo, T. Highly redundant function of multiple AT-rich sequences as core promoter elements in the TATA-less RPS5 promoter of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **39**, 59–75 (2011).
11. Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP *in vivo* and *in vitro* reveals a direct functional interaction with the SAGA subunit Spt3. *Genes Dev.* **22**, 2994–3006 (2008).
12. Dudley, A. M., Rougeulle, C. & Winston, F. The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step *in vivo*. *Genes Dev.* **13**, 2940–2945 (1999).
13. Bhaumik, S. R. & Green, M. R. Differential requirement of SAGA components for recruitment of TATA-box-binding protein to promoters *in vivo*. *Mol. Cell. Biol.* **22**, 7365–7371 (2002).
14. Kostrewa, D. *et al.* RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* **462**, 323–330 (2009).
15. Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* **327**, 206–209 (2010).
16. Li, Y., Flanagan, P. M., Tschochner, H. & Kornberg, R. D. RNA polymerase II initiation factor interactions and transcription start site selection. *Science* **263**, 805–807 (1994).
17. Pardee, T. S., Bangur, C. S. & Ponticelli, A. S. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J. Biol. Chem.* **273**, 17859–17864 (1998).
18. Yan, Q., Moreland, R. J., Conaway, J. W. & Conaway, R. C. Dual roles for transcription factor IIIf in promoter escape by RNA polymerase II. *J. Biol. Chem.* **274**, 35668–35675 (1999).
19. Kim, T. K., Ebright, R. H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIIf. *Science* **288**, 1418–1421 (2000).
20. Keogh, M. C., Cho, E. J., Podolny, V. & Buratowski, S. Kin28 is found within TFIIf and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. *Mol. Cell. Biol.* **22**, 1288–1297 (2002).
21. Sveistrup, J. Q., Feaver, W. J. & Kornberg, R. D. Subunits of yeast RNA polymerase II transcription factor TFIIf encoded by the *CCL1* gene. *J. Biol. Chem.* **271**, 643–645 (1996).
22. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein–DNA interactions detected at single nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
23. Giardina, C. & Lis, J. T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
24. Giardina, C. & Lis, J. T. Dynamic protein–DNA architecture of a yeast heat shock promoter. *Mol. Cell. Biol.* **15**, 2737–2744 (1995).
25. Forget, D., Langlier, M. F., Therien, C., Trinh, V. & Coulombe, B. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol. Cell. Biol.* **24**, 1122–1131 (2004).
26. Lagrange, T. *et al.* High-resolution mapping of nucleoprotein complexes by site-specific protein–DNA photocrosslinking: organization of the human TBP-TFIIf-TFIIB–DNA quaternary complex. *Proc. Natl. Acad. Sci. USA* **93**, 10620–10625 (1996).
27. Chen, H. T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**, 169–180 (2004).
28. Pal, M., Ponticelli, A. S. & Luse, D. S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005).

29. Kuehner, J. N. & Brow, D. A. Quantitative analysis of *in vivo* initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.* **281**, 14119–14128 (2006).
30. Matangkasombut, O., Buratowski, R. M., Swilling, N. W. & Buratowski, S. Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes Dev.* **14**, 951–962 (2000).
31. Koerber, R. T., Rhee, H. S., Jiang, C. & Pugh, B. F. Interaction of transcriptional regulators with specific nucleosomes across the *Saccharomyces* genome. *Mol. Cell.* **35**, 889–902 (2009).
32. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
33. Faitar, S. L., Brodie, S. A. & Ponticelli, A. S. Promoter-specific shifts in transcription initiation conferred by yeast TFIIB mutations are determined by the sequence in the immediate vicinity of the start sites. *Mol. Cell. Biol.* **21**, 4427–4440 (2001).
34. Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nature Struct. Mol. Biol.* **17**, 1272–1278 (2010).
35. Ahn, S. H., Keogh, M. C. & Buratowski, S. Ctk1 promotes dissociation of basal transcription factors from elongating RNA polymerase II. *EMBO J.* **28**, 205–212 (2009).
36. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Struct. Mol. Biol.* **11**, 394–403 (2004).
37. Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).
38. Flom, G., Weekes, J. & Johnson, J. L. Novel interaction of the Hsp90 chaperone machine with Ssl2, an essential DNA helicase in *Saccharomyces cerevisiae*. *Curr. Genet.* **47**, 368–380 (2005).
39. Freeman, B. C. & Yamamoto, K. R. Disassembly of transcriptional regulatory complexes by molecular chaperones. *Science* **296**, 2232–2235 (2002).
40. Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10**, 833–844 (2009).
41. Wei, W., Pelechano, V., Jarvelin, A. I. & Steinmetz, L. M. Functional consequences of bidirectional promoters. *Trends Genet.* **27**, 267–276 (2011).
42. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
43. Granovskaia, M. V. *et al.* High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol.* **11**, R24 (2010).
44. van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114–117 (2011).
45. Gerstein, M. B. *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17**, 669–681 (2007).
46. Albert, I., Wachi, S., Jiang, C. & Pugh, B. F. GeneTrack—a genomic data processing and visualization framework. *Bioinformatics* **24**, 1305–1306 (2008).
47. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* **103**, 5320–5325 (2006).
48. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
49. Ansari, A. & Hampsey, M. A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev.* **19**, 2969–2978 (2005).
50. Singh, B. N. & Hampsey, M. A transcription-independent role for TFIIB in gene looping. *Mol. Cell.* **27**, 806–816 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Albert and Y. Li for bioinformatic support, and members of the Pugh laboratory and the Penn State Center for Eukaryotic Gene Regulation for valuable discussions. Sequencing was performed at the Penn State Genomics Core Facility. This work was supported by National Institutes of Health grant GM059055.

Author Contributions H.S.R. performed the experiments and conducted data analyses. H.S.R. and B.F.P. conceived the experiments, analyses, and co-wrote the manuscript.

Author Information Sequencing data have been deposited at the NCBI Sequence Read Archive under accession number SRA046523. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.F.P. (bfp@psu.edu).

Creating, moving and merging Dirac points with a Fermi gas in a tunable honeycomb lattice

Leticia Tarruell¹, Daniel Greif¹, Thomas Uehlinger¹, Gregor Jotzu¹ & Tilman Esslinger¹

Dirac points are central to many phenomena in condensed-matter physics, from massless electrons in graphene to the emergence of conducting edge states in topological insulators^{1,2}. At a Dirac point, two energy bands intersect linearly and the electrons behave as relativistic Dirac fermions. In solids, the rigid structure of the material determines the mass and velocity of the electrons, as well as their interactions. A different, highly flexible means of studying condensed-matter phenomena is to create model systems using ultracold atoms trapped in the periodic potential of interfering laser beams^{3,4}. Here we report the creation of Dirac points with adjustable properties in a tunable honeycomb optical lattice. Using momentum-resolved interband transitions, we observe a minimum bandgap inside the Brillouin zone at the positions of the two Dirac points. We exploit the unique tunability of our lattice potential to adjust the effective mass of the Dirac fermions by breaking inversion symmetry. Moreover, changing the lattice anisotropy allows us to change the positions of the Dirac points inside the Brillouin zone. When the anisotropy exceeds a critical limit, the two Dirac points merge and annihilate each other—a situation that has recently attracted considerable theoretical interest^{5–9} but that is extremely challenging to observe in solids¹⁰. We map out this topological transition in lattice parameter space and find excellent agreement with *ab initio* calculations. Our results not only pave the way to model materials in which the topology of the band structure is crucial, but also provide an avenue to exploring many-body phases resulting from the interplay of complex lattice geometries with interactions^{11–13}.

Ultracold gases in optical lattices have become a versatile tool with which to simulate a wide range of condensed-matter phenomena^{3,4}. For example, the control of interactions has led to the observation of Mott insulating phases^{14–16}. In fermionic systems, this provides new access to the physics of strongly correlated materials. However, the topology of the band structure is equally important for the properties of a solid. A prime example is the honeycomb lattice of graphene, where the presence of topological defects in momentum space—the Dirac points—leads to remarkable transport properties, even in the absence of interactions¹. In quantum gases, a honeycomb lattice has recently been realized and investigated using a Bose–Einstein condensate^{17,18}, but no signatures of Dirac points were observed. Here we study an ultracold Fermi gas of ⁴⁰K atoms in a two-dimensional, tunable optical lattice, which can be continuously adjusted to create square, triangular, dimer and honeycomb structures. In the honeycomb lattice, we identify the presence of Dirac points in the band structure by observing a minimum bandgap inside the Brillouin zone using interband transitions. Our method is closely related to a technique recently used with bosonic atoms to characterize the linear crossing of two high-energy bands in a one-dimensional, bichromatic lattice¹⁹, but also provides momentum resolution.

To create and manipulate Dirac points, we have developed a two-dimensional optical lattice of adjustable geometry. It is formed by three retro-reflected laser beams of wavelength $\lambda = 1,064$ nm, arranged as depicted in Fig. 1a. The interference of two perpendicular beams, X

and Y, gives rise to a chequerboard lattice of spacing $\lambda/\sqrt{2}$. A third beam, \bar{X} , collinear with X but detuned by a frequency δ , creates an additional standing wave with a spacing of $\lambda/2$. This yields a potential of the form

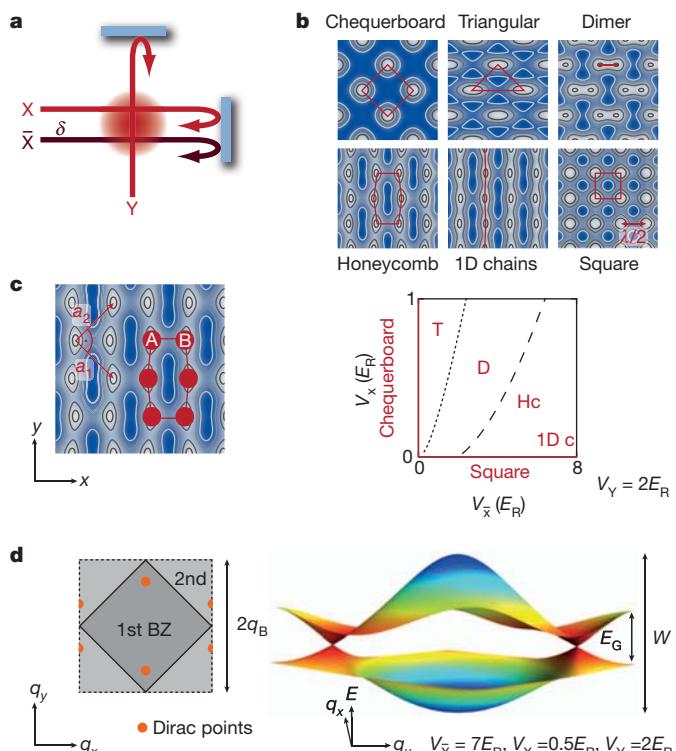


Figure 1 | Optical lattice with adjustable geometry. **a**, Three retro-reflected laser beams of wavelength $\lambda = 1,064$ nm create the two-dimensional lattice potential of equation (1). Beams X and Y interfere and produce a chequerboard pattern, and beam \bar{X} creates an independent standing wave. Their relative position is controlled by the detuning δ . **b**, Top: different lattice potentials can be realized depending on the intensities of the lattice beams. White regions correspond to lower potential energies and blue regions to higher potential energies. Bottom: diagram showing the accessible lattice geometries as a function of the lattice depths $V_{\bar{X}}$ and V_X . The transition between triangular (T) and dimer (D) lattices is indicated by a dotted line. When crossing the dashed line into the honeycomb (Hc) regime, Dirac points appear. The limit $V_X \gg V_{\bar{X}}$, $V_{\bar{X}} \gg V_Y$ corresponds to weakly coupled, one-dimensional chains (1D c). **c**, The real-space potential of the honeycomb lattice has a two-site unit cell (sites A and B) and the primitive lattice vectors are perpendicular. **d**, Left: sketch of the first and second Brillouin zones (BZs) of the honeycomb lattice, indicating the positions of the Dirac points. Right: three-dimensional view of the energy spectrum showing the linear intersection of the bands at the two Dirac points. The colour scale illustrates lines of constant energy. We denote the full bandwidth, W ; the minimum energy gap at the edges of the Brillouin zone, E_G ; and the Bloch wavevector, $q_B = 2\pi/\lambda$.

¹Institute for Quantum Electronics, ETH Zurich, 8093 Zurich, Switzerland.

$$V(x, y) = -V_{\bar{X}} \cos^2(kx + \theta/2) - V_X \cos^2(kx) - V_Y \cos^2(ky) - 2\alpha \sqrt{V_X V_Y} \cos(kx) \cos(ky) \cos(\varphi) \quad (1)$$

where $V_{\bar{X}}$, V_X and V_Y denote the single-beam lattice depths (proportional to the laser beam intensities), α is the visibility of the interference pattern and $k = 2\pi/\lambda$. We can adjust the two phases continuously, and choose $\theta = \pi$ and $\varphi = 0$ (Methods). Varying the relative intensities of the beams allows us to realize various lattice structures (Fig. 1b). In the following, we focus on the honeycomb lattice, whose real-space potential is shown in Fig. 1c.

The honeycomb lattice consists of two sublattices, A and B. Therefore, the wavefunctions are two-component spinors¹. Tunnelling between the sublattices leads to the formation of two energy bands, which are well separated from the higher bands and have a conical intersection at two quasi-momentum points in the Brillouin zone—the Dirac points. These points are topological defects in the band structure, with respective associated Berry phases of π and $-\pi$. This guarantees their stability with respect to lattice perturbations, such that a large range of lattice anisotropies change only the positions of the Dirac points inside the Brillouin zone. In contrast, breaking the inversion symmetry of the potential by introducing an energy offset, Δ , between the sublattices opens an energy gap at the Dirac points, proportional to Δ . In our implementation, Δ depends only on the value of the phase θ and can be precisely adjusted (Methods). As shown in Fig. 1c, d, the primitive lattice vectors are perpendicular, leading to a square Brillouin zone with two Dirac points inside. Their positions are symmetric around the centre and are fixed to quasi-momentum $q_x = 0$, owing to the time-reversal and reflection symmetries of the system²⁰. The band structure for our lattice implementation is in the two lowest bands topologically equivalent to that of a hexagonal lattice with six-fold symmetry. For deep lattices, both configurations then also map to the same tight-binding Hamiltonian.

We characterize the Dirac points by probing the energy splitting between the two lowest-energy bands through interband transitions. The starting point of the experiment is a non-interacting, ultracold gas of $N \approx 50,000$ fermionic ^{40}K atoms in the $|F, m_F\rangle = |9/2, -9/2\rangle$ state, where F denotes the hyperfine manifold and m_F the Zeeman state. The cloud is prepared in the lowest-energy band of a honeycomb lattice with $V_{\bar{X}}/E_R = 4.0(2)$, $V_X/E_R = 0.28(1)$ and $V_Y/E_R = 1.8(1)$, which also causes a weak harmonic confinement with trapping frequencies $\omega_x/2\pi = 17.6(1)$ Hz, $\omega_y/2\pi = 31.8(5)$ Hz and $\omega_z/2\pi = 32.7(5)$ Hz. Here $E_R = \hbar^2/2m\lambda^2$ is the recoil energy, \hbar denotes Planck's constant and m is the mass of a ^{40}K atom. Throughout the manuscript, errors in parenthesis denote the standard deviation. On application of a weak magnetic field gradient, the atomic cloud is subjected to a constant force, F , in the x direction, with an effect equivalent to that produced by an electric field in solid-state systems. The atoms are hence accelerated such that their quasi-momentum q_x increases linearly up to the edge of the Brillouin zone, where a Bragg reflection occurs. The cloud eventually returns to the centre of the band, performing one full Bloch oscillation²¹. We then measure the quasi-momentum distribution of the atoms in the different bands²² (Methods).

Owing to the finite momentum width of the cloud, trajectories with different quasi-momenta q_y are simultaneously explored during the Bloch cycle (Fig. 2a). For a trajectory far from the Dirac points, the atoms remain in the lowest-energy band (trajectory 1). In contrast, when passing through a Dirac point (trajectory 2), the atoms are transferred from the first band to the second because of the vanishing energy splitting at the linear band crossing. When measuring the quasi-momentum distribution, these atoms are missing in the first Brillouin zone and appear in the second band (Fig. 2a). We identify the points of maximum transfer with the Dirac points. The energy resolution of the method is set by the characteristic energy of the applied force²¹, $E_B/h = F\lambda/2h = 88.6(7)$ Hz, which is small compared

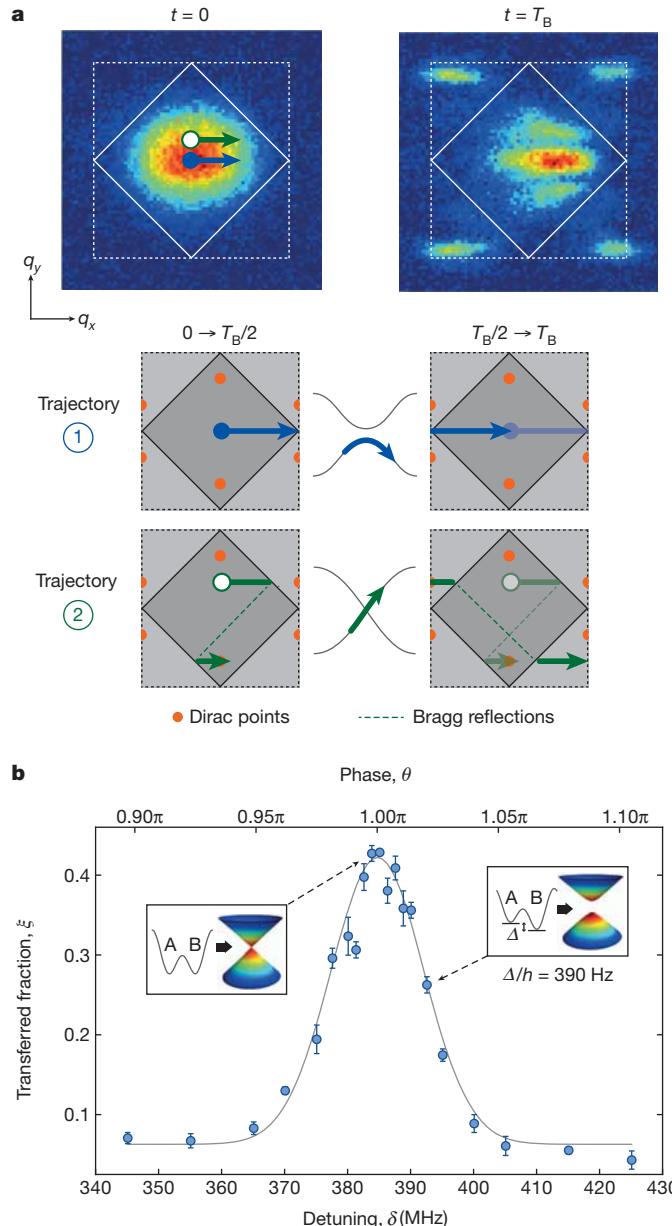


Figure 2 | Probing the Dirac points. **a**, Quasi-momentum distribution of the atoms before and after one Bloch oscillation, of period T_B (colour scale, column density of the absorption image in arbitrary units). The cloud explores several trajectories in quasi-momentum space simultaneously. For trajectory 1 (blue filled circle), the atoms remain in the first energy band. In contrast, trajectory 2 (green open circle) passes through a Dirac point at $t = T_B/2$. There the energy splitting between the bands vanishes and the atoms are transferred to the second band. When measuring the quasi-momentum distribution at $t = T_B$, these atoms are missing from the first Brillouin zone and appear in the second one. **b**, Dependence of the total fraction of atoms transferred to the second band, ξ , on the detuning, δ , of the lattice beams, which controls the sublattice energy offset, Δ . The maximum indicates the point of inversion symmetry, where $\Delta = 0$ ($\theta = \pi$ in equation (1)) and the gap at the Dirac point vanishes. Insets: away from the peak, the atoms behave as Dirac fermions with a tunable mass. Data show mean \pm s.d. of five consecutive measurements; solid line is a Gaussian fit to the data.

with the full bandwidth, $W/h = 4.6$ kHz, and the minimum bandgap at the edges of the Brillouin zone, $E_G/h = 475$ Hz.

To investigate how breaking the inversion symmetry of the lattice affects the Dirac points, we vary the sublattice offset, Δ , which is controlled by the frequency detuning, δ , between the lattice beams, and measure the total fraction of atoms transferred to the second band,

ξ . The results obtained for a honeycomb lattice with $V_{\bar{X}}/E_R = 3.6(2)$, $V_X/E_R = 0.28(1)$ and $V_Y/E_R = 1.8(1)$ are displayed in Fig. 2b and show a sharp maximum in the transferred fraction. We identify this situation as the point of inversion symmetry, where $\Delta = 0$ ($\theta = \pi$), in good agreement with an independent calibration (Methods). At this setting, the bandgap at the Dirac points vanishes. The population in the second band decreases symmetrically on both sides of the peak as the gap increases, indicating the transition from massless to massive Dirac fermions.

The relative strength of the tunnel couplings between the different sites of the lattice fixes the position of the Dirac points inside the Brillouin zone, as well as the slope of the associated linear dispersion relation^{5–9}. However, the tunability of our optical lattice structure allows for independent adjustment of the tunnelling parameters in the x and y directions simply by controlling the intensity of the laser beams. For isotropic tunnellings, the slope of the dispersion relation around the Dirac points is the same in all directions, but is anisotropic otherwise. The distance from the Dirac points to the corners of the Brillouin zone along q_y can be varied between 0 and $q_B/2$, whereas $q_x = 0$ is fixed by reflection symmetry²⁰. Here $q_B = 2\pi/\lambda$ denotes the Bloch wave vector.

We exploit the momentum resolution of the interband transitions directly to observe the movement of the Dirac points. Starting from a honeycomb lattice with $V_{\bar{X}}/E_R = 5.4(3)$, $V_X/E_R = 0.28(1)$ and $V_Y/E_R = 1.8(1)$, we gradually increase the tunnelling in the x direction by decreasing the intensity of \bar{X} . The position of the Dirac points continuously approaches the corners of the Brillouin zone (Fig. 3), as expected from an *ab initio* two-dimensional band structure calculation (Methods). The deviations close to the merging point are possibly

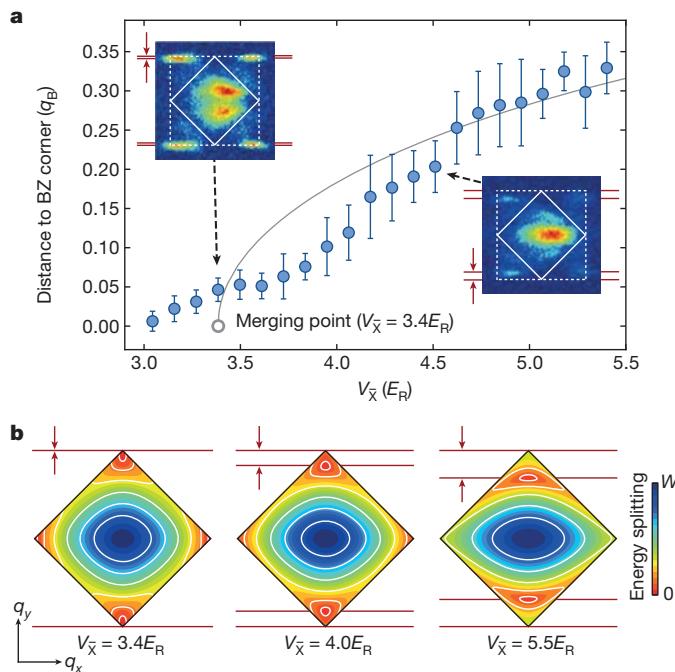


Figure 3 | Movement of the Dirac points. **a**, Distance from the Dirac points to the corners of the Brillouin zone, as measured through momentum-resolved interband transitions. The tunnelling in the x direction increases when the lattice depth $V_{\bar{X}}$ is decreased. The distance is extracted from the second-band quasi-momentum distribution after one Bloch cycle (insets). The merging of the two Dirac points at the corners of the Brillouin zone is signalled by a single line of missing atoms in the first band. Data show mean \pm s.d. of three to nine measurements. The solid line is the prediction of a two-dimensional band structure calculation without any fitting parameters. **b**, Energy splitting between the two lowest bands. It shows the displacement of the Dirac cones inside the Brillouin zone, as well as their deformation depending on the lattice depth V_X .

caused by the flattening of the dispersion relation between the two Dirac points as they approach each other⁸.

When they reach the corners of the Brillouin zone, the two Dirac points merge, annihilating each other. There the dispersion relation becomes quadratic along the q_y axis, remaining linear along q_x . Beyond this critical point, a finite bandgap appears for all quasi-momenta of the Brillouin zone. This situation signals the transition between band structures of two different topologies, one containing two Dirac points and the other containing none. For two-dimensional honeycomb lattices at half-filling, it corresponds to a Lifshitz phase transition from a semimetallic phase to a band-insulating phase^{6,7}.

We experimentally map out the topological transition line by recording the fraction of atoms transferred to the second band, ξ , as a function of the lattice depths $V_{\bar{X}}$ and V_X , while keeping $V_Y/E_R = 1.8(1)$. The results are shown in Fig. 4a. There the onset of population transfer to the second band signals the appearance of Dirac points in the band structure of the lattice. For a given value of V_X , the transferred fraction, ξ , decreases again for large values of $V_{\bar{X}}$, as the Dirac points lie beyond the momentum width of the cloud.

To extend the range of our measurements and probe the Dirac points even in this region, we apply a force in the y direction. We hence explore a new class of trajectories in quasi-momentum space. This allows for the investigation of very anisotropic Dirac cones, which become almost flat in the q_x direction as we approach the crossover to a one-dimensional lattice structure ($V_{\bar{X}} \gg V_X$). Along the q_y trajectories, the centre of the cloud successively passes the two Dirac points during the Bloch cycle, effectively realizing a Stückelberg interferometer^{23,24} in a two-dimensional band structure. As shown in Fig. 4b, we again identify the topological transition by the onset of population transfer to the second band. The results for the transition lines obtained for the

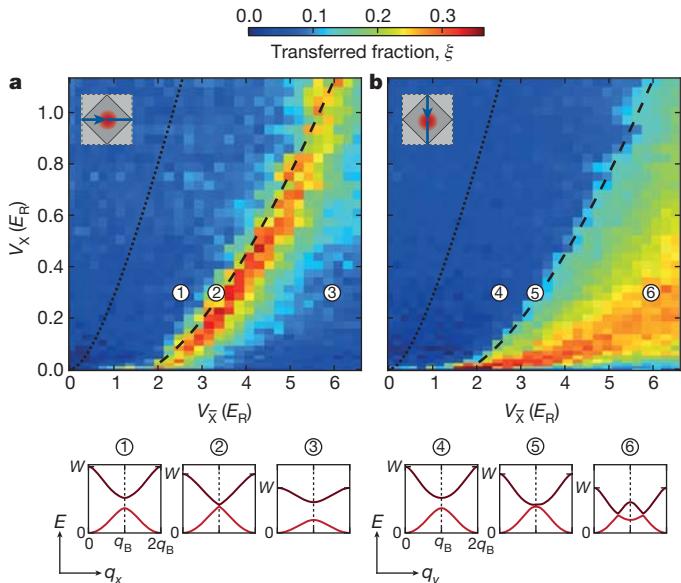


Figure 4 | Topological transition. Fraction of atoms transferred to the second band, ξ , as a function of lattice depths $V_{\bar{X}}$ and V_X , with $V_Y/E_R = 1.8(1)$. Different lattice geometries (square, chequerboard, triangular, dimer and honeycomb) are realized (Fig. 1b). We consider trajectories in quasi-momentum space in the q_x (a) and q_y (b) directions. Each data point is a single measurement, and as a result there are at least 1,200 points per diagram. To maximize the transfer for the q_y trajectories, where the cloud successively passes the two Dirac points, we set $\theta = 1.013(1)\pi$. For both trajectories, the onset of population transfer to the second band signals the topological transition, where the Dirac points appear. The dashed line is the theoretical prediction for the transition line without any fitting parameters, and the dotted line indicates the transition from the triangular lattice to the dimer lattice. The bottom diagrams show cuts of the band structure along the q_x axis ($q_y = 0$; a) and q_y axis ($q_x = 0$; b) for the values of V_X and $V_{\bar{X}}$ indicated.

two measurement series are in excellent agreement with an *ab initio* band structure calculation. At the transition, the transfer is expected to be lower by a factor of two along the central q_y trajectory, as compared to the central q_x trajectory. This is caused by the dispersion relation at the corners of the Brillouin zone being quadratic in the q_y direction and linear in q_x . For the data points on the (dashed) transition line in Fig. 4, we find an average ratio for the transfer along q_y and q_x of $\xi = 0.52(14)$, in good agreement with the simple model.

In this work, we have realized Dirac points with highly tunable properties using ultracold fermionic atoms in a honeycomb optical lattice. A new class of physical phenomena based on complex lattice topologies is now within the domain of quantum gas experiments. The versatility and novel observables of these systems will provide new insights. For example, the unique coherence of quantum gases offers the possibility of directly measuring, by interferometric methods, the Berry phase associated with the Dirac points. Topological order could be achieved by introducing artificial gauge fields, either through Raman transitions²⁵ or time-dependent lattice modulation²⁶. Moreover, the highly tunable lattice potential we have developed opens many new avenues for optical lattice experiments. For spin mixtures with repulsive interactions, the dynamic transition between dimer and square lattices should facilitate the adiabatic preparation of an antiferromagnetic phase²⁷ and allow the study of quantum criticality²⁸. Additionally, the triangular and honeycomb lattices provide the possibility of exploring magnetic frustration and spin liquid phases^{12,13}.

METHODS SUMMARY

To prepare a non-interacting, quantum-degenerate Fermi gas of ^{40}K atoms, we perform evaporative cooling of a balanced spin mixture of the $|F, m_F\rangle = |9/2, -9/2\rangle$ and $|9/2, -7/2\rangle$ states in an optical dipole trap at a magnetic field of $197.6(1)\text{ G}$. After reaching typical temperatures of $0.2T_F$, where T_F is the Fermi temperature, we apply a magnetic field gradient in the y direction to remove the $|9/2, -7/2\rangle$ component while levitating the $|9/2, -9/2\rangle$ atoms. The polarized Fermi gas is loaded into the optical lattice, where the dipole trap is switched off. We then apply an additional magnetic field gradient in either the x or the y direction, to excite Bloch oscillations. The two-dimensional lattice lies in the x - y plane, whereas in the z direction the atoms are harmonically trapped. The time phase, φ , between the interfering beams X and Y in equation (1) is stabilized interferometrically at a value of $\varphi/\pi = 0.00(3)$. The measured visibility of the interference pattern is $\alpha = 0.90(5)$. The phase θ depends on the optical path length between the atomic cloud and the retro-reflecting mirror, and can be precisely controlled through the detuning, δ , between beams \bar{X} and X. We infer the precise value $\theta/\delta = \pi/384.7(6)\text{ MHz}^{-1}$ from the peak position in Fig. 2b. This is in good agreement with an independent calibration obtained using Raman–Nath diffraction on a ^{87}Rb Bose–Einstein condensate, which yields $\theta/\delta = \pi/388(4)\text{ MHz}^{-1}$. The theory curves are extracted from an *ab initio* single-particle, two-dimensional, numerical band structure calculation for the homogeneous system. It therefore also takes into account higher-order tunnelling terms, which are relevant for the regime studied here. In particular, they cause an asymmetry between the two lowest bands and lead to a tilt of the Dirac cones in certain parameter regimes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 November 2011; accepted 17 January 2012.

1. Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109–162 (2009).

2. Hasan, M. Z. & Kane, C. L. Topological insulators. *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
3. Lewenstein, M. et al. Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond. *Adv. Phys.* **56**, 243–379 (2007).
4. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).
5. Hasegawa, Y., Konno, R., Nakano, H. & Kohmoto, M. Zero modes of tight-binding electrons on the honeycomb lattice. *Phys. Rev. B* **74**, 033413 (2006).
6. Zhu, S.-L., Wang, B. & Duan, L.-M. Simulation and detection of Dirac fermions with cold atoms in an optical lattice. *Phys. Rev. Lett.* **98**, 260402 (2007).
7. Wunsch, B., Guinea, F. & Sols, F. Dirac-point engineering and topological phase transitions in honeycomb optical lattices. *N. J. Phys.* **10**, 103027 (2008).
8. Montambaux, G., Piéchon, F., Fuchs, J.-N. & Goerbig, M. O. Merging of Dirac points in a two-dimensional crystal. *Phys. Rev. B* **80**, 153412 (2009).
9. Lee, K. L., Grémaud, B., Han, R., Englert, B.-G. & Miniatura, C. Ultracold fermions in a graphene-type optical lattice. *Phys. Rev. A* **80**, 043411 (2009).
10. Pereira, V. M., Castro Neto, A. H. & Peres, N. M. R. Tight-binding approach to uniaxial strain in graphene. *Phys. Rev. B* **80**, 045401 (2009).
11. Zhao, E. & Parameswari, A. BCS–BEC crossover on the two-dimensional honeycomb lattice. *Phys. Rev. Lett.* **97**, 230404 (2006).
12. Balents, L. Spin liquids in frustrated magnets. *Nature* **464**, 199–208 (2010).
13. Meng, Z. Y., Lang, T. C., Wessel, S., Assaad, F. F. & Muramatsu, A. Quantum spin liquid emerging in two-dimensional correlated Dirac fermions. *Nature* **464**, 847–851 (2010).
14. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415**, 39–44 (2002).
15. Jördens, R., Strohmaier, N., Günter, K., Moritz, H. & Esslinger, T. A Mott insulator of fermionic atoms in an optical lattice. *Nature* **455**, 204–207 (2008).
16. Schneider, U. et al. Metallic and insulating phases of repulsively interacting fermions in a 3D optical lattice. *Science* **322**, 1520–1525 (2008).
17. Soltan-Panahi, P. et al. Multi-component quantum gases in spin-dependent hexagonal lattices. *Nature Phys.* **7**, 434–440 (2011).
18. Soltan-Panahi, P., Lühmann, D.-S., Struck, J., Windpassinger, P. & Sengstock, K. Quantum phase transition to unconventional multi-orbital superfluidity in optical lattices. *Nature Phys.* **8**, 71–75 (2012).
19. Salger, T., Geckeler, C., Kling, S. & Weitz, M. Atomic Landau-Zener tunneling in Fourier-synthesized optical lattices. *Phys. Rev. Lett.* **99**, 190405 (2007).
20. Asano, K. & Hotta, C. Designing Dirac points in two-dimensional lattices. *Phys. Rev. B* **83**, 245125 (2011).
21. Ben Dahan, M., Peik, E., Reichel, J., Castin, Y. & Salomon, C. Bloch oscillations of atoms in an optical potential. *Phys. Rev. Lett.* **76**, 4508–4511 (1996).
22. Köhl, M., Moritz, H., Stöferle, T., Günter, K. & Esslinger, T. Fermionic atoms in a three dimensional optical lattice: observing Fermi surfaces, dynamics, and interactions. *Phys. Rev. Lett.* **94**, 080403 (2005).
23. Kling, S., Salger, T., Grossert, C. & Weitz, M. Atomic Bloch-Zener oscillations and Stückelberg interferometry in optical lattices. *Phys. Rev. Lett.* **105**, 215301 (2010).
24. Zenesini, A., Ciampini, D., Morsch, O. & Arimondo, E. Observation of Stückelberg oscillations in accelerated optical lattices. *Phys. Rev. A* **82**, 065601 (2010).
25. Lin, Y.-J., Compton, R. L., Jiménez-García, K., Porto, J. V. & Spielman, I. B. Synthetic magnetic fields for ultracold neutral atoms. *Nature* **462**, 628–632 (2009).
26. Kitagawa, T., Berg, E., Rudner, M. & Demler, E. Topological characterization of periodically driven quantum systems. *Phys. Rev. B* **82**, 235114 (2010).
27. Lubasch, M., Murg, V., Schneider, U., Cirac, J. I. & Bañuls, M.-C. Adiabatic preparation of a Heisenberg antiferromagnet using an optical superlattice. *Phys. Rev. Lett.* **107**, 165301 (2011).
28. Sachdev, S. Quantum magnetism and criticality. *Nature Phys.* **4**, 173–185 (2008).

Acknowledgements We would like to thank D. Poletti for bringing our attention to honeycomb lattices without six-fold symmetry, and N. Cooper and F. Hassler for discussions. We acknowledge SNF, NCCR-MaNEP, NCCR-QSIT, NAME-QUM (EU, FET open), SQMS (ERC advanced grant) and ESF (POLATOM) for funding.

Author Contributions The data were measured and analysed by L.T., D.G., T.U. and G.J. The tunable optical lattice was built by D.G. The experimental concept was developed by T.E. All authors contributed extensively to the discussion of the results, as well as to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.E. (esslinger@phys.ethz.ch).

METHODS

Preparation scheme. After sympathetic cooling with ^{87}Rb in a magnetic trap, 2×10^6 fermionic ^{40}K atoms are transferred into a dipole trap operating at a wavelength of 826 nm. A balanced spin mixture of the $|F, m_F\rangle = |9/2, -9/2\rangle$ and $|9/2, -7/2\rangle$ states is evaporatively cooled in a magnetic field of 197.6(1) G, in the vicinity of the Feshbach resonance at 202.1 G. We obtain typical temperatures of $0.2T_F$. The field is subsequently reduced and a magnetic field gradient is used to remove the $|9/2, -7/2\rangle$ component while levitating the $|9/2, -9/2\rangle$ atoms. This polarized Fermi gas is loaded into the two-dimensional optical lattice in 200 ms, and the dipole trap is then switched off.

Tunable optical lattice. The optical lattice is produced by the combination of three retro-reflected laser beams of linear polarization and wavelength $\lambda = 1,064$ nm, red-detuned with respect to the D_1 and D_2 lines of ^{40}K . Beams \bar{X} and X propagate along the same axis and beam Y propagates at $90.0(1)^\circ$ to that axis. The interference of X and Y produces a chequerboard potential of the form²⁹

$$V_1(x, y) = -V_X \cos^2(kx) - V_Y \cos^2(ky) - 2\alpha\sqrt{V_X V_Y} \cos(kx) \cos(ky) \cos(\phi)$$

The phase between the two beams at the position of the atoms is stabilized interferometrically at a value of $\phi/\pi = 0.00(3)$ using a pair of additional beams detuned from each other and from \bar{X} , X and Y . This results in a weak additional lattice along each axis of about $0.1E_R$. We use a heterodyne locking technique³⁰, where the two interfering arms are formed by the optical lattice beam paths and are each compared to a common reference beam.

Beam \bar{X} creates a potential $V_2(x) = -V_{\bar{X}} \cos^2(kx + \theta/2)$, where the phase θ determines the relative position of the chequerboard pattern and the one-dimensional standing wave. We control the value of θ in the centre of the cloud by adjusting the frequency detuning, δ , between \bar{X} and X . We infer the precise value $\theta/\delta = \pi/384.7(6)$ MHz⁻¹ from the peak position in Fig. 2b. This is in good agreement with an independent calibration obtained using Raman–Nath diffraction on a ^{87}Rb Bose–Einstein condensate, which yields $\theta/\delta = \pi/388(4)$ MHz⁻¹. At the edges of the cloud, the phase differs by approximately $\pm 10^{-4}\pi$.

The total lattice potential is given by $V_1(x, y) + V_2(x)$ and, depending on the relative intensities of the beams, gives rise to square, chequerboard, triangular and honeycomb lattices, as well as to a staggered arrangement of dimers and an array of weakly coupled, one-dimensional chains.

The visibility, $\alpha = 0.90(5)$, and the lattice depths, $V_{\bar{X}}$, V_X and V_Y , are calibrated using Raman–Nath diffraction. The method has a systematic uncertainty of 10% for the lattice depths; the statistical uncertainties are given in the main text. The two-dimensional lattice lies in the x – y plane, whereas in the z direction the atoms are harmonically trapped. Owing to the absence of interactions, the z direction decouples.

The underlying trap frequencies in our system scale with the lattice depths according to the approximate expressions $\omega_x \propto \sqrt{V_Y}$, $\omega_y \propto \sqrt{V_{\bar{X}} + V_X V_Y / V_{\bar{X}}}$ and $\omega_z \propto \sqrt{V_{\bar{X}} + V_X V_Y / V_{\bar{X}}} + 1.24V_Y$. For the parameters used in Fig. 2a, we find that $\omega_x/2\pi = 17.6(1)$ Hz, $\omega_y/2\pi = 31.8(5)$ Hz and $\omega_z/2\pi = 32.7(5)$ Hz, as calibrated from dipole oscillations of the cloud.

Detection. The quasi-momentum distribution of the gas is probed using a band-mapping technique. The optical lattice beams are linearly ramped down in 500 μs , that is, slowly enough for the atoms to stay adiabatically in their band while quasi-momentum is approximately conserved²². We then allow for 15 ms of ballistic expansion before making an absorption image of the cloud.

Band structure calculations. The energy spectrum is obtained using an *ab initio* single-particle, two-dimensional, numerical band structure calculation for the homogeneous system. It therefore also takes into account higher-order tunnelling terms, which are relevant for the regime studied here. In particular, they cause an asymmetry between the two lowest bands and lead to a tilt of the Dirac cones in certain parameter regimes.

29. Hemmerich, A., Schropp, D., Esslinger, T. & Hänsch, T. W. Elastic scattering of rubidium atoms by two crossed standing waves. *Europhys. Lett.* **18**, 391–395 (1992).
30. Ma, L.-S., Jungner, P., Ye, J. & Hall, J. L. Delivering the same optical frequency at two places: accurate cancellation of phase noise introduced by an optical fiber or other time-varying path. *Opt. Lett.* **19**, 1777–1779 (1994).

Designer Dirac fermions and topological phases in molecular graphene

Kenjiro K. Gomes^{1,2*}, Warren Mar^{2,3*}, Wonhee Ko^{2,4*}, Francisco Guinea⁵ & Hari C. Manoharan^{1,2}

The observation of massless Dirac fermions in monolayer graphene has generated a new area of science and technology seeking to harness charge carriers that behave relativistically within solid-state materials¹. Both massless and massive Dirac fermions have been studied and proposed in a growing class of Dirac materials that includes bilayer graphene, surface states of topological insulators and iron-based high-temperature superconductors. Because the accessibility of this physics is predicated on the synthesis of new materials, the quest for Dirac quasi-particles has expanded to artificial systems such as lattices comprising ultracold atoms^{2–4}. Here we report the emergence of Dirac fermions in a fully tunable condensed-matter system—molecular graphene—assembled by atomic manipulation of carbon monoxide molecules over a conventional two-dimensional electron system at a copper surface⁵. Using low-temperature scanning tunnelling microscopy and spectroscopy, we embed the symmetries underlying the two-dimensional Dirac equation into electron lattices, and then visualize and shape the resulting ground states. These experiments show the existence within the system of linearly dispersing, massless quasi-particles accompanied by a density of states characteristic of graphene. We then tune the quantum tunnelling between lattice sites locally to adjust the phase accrual of propagating electrons. Spatial texturing of lattice distortions produces atomically sharp p–n and p–n–p junction devices with two-dimensional control of Dirac fermion density and the power to endow Dirac particles with mass^{6–8}. Moreover, we apply scalar and vector potentials locally and globally to engender topologically distinct ground states and, ultimately, embedded gauge fields^{9–12}, wherein Dirac electrons react to ‘pseudo’ electric and magnetic fields present in their reference frame but absent from the laboratory frame. We demonstrate that Landau levels created by these gauge fields can be taken to the relativistic magnetic quantum limit, which has so far been inaccessible in natural graphene. Molecular graphene provides a versatile means of synthesizing exotic topological electronic phases in condensed matter using tailored nanostructures.

The Dirac fermion has emerged as a common feature of new materials whose band structure and embedded spin degree of freedom are described by the relativistic Dirac equation¹. In two dimensions, it has long been appreciated theoretically that the honeycomb lattice and mapping of the sublattice degree of freedom to a pseudospin is represented by the Dirac equation for electrons bound to the lattice, $H_G = \hbar\tilde{c}\sigma \cdot \mathbf{k}$, where \mathbf{k} is the vector momentum, $\sigma = (\sigma_x, \sigma_y)$ and σ_i are the 2×2 Pauli spin matrices coupled to pseudospin, \tilde{c} is the Dirac fermion velocity (the effective speed of light) and \hbar is Planck’s constant divided by 2π . The simplest model for this physics is captured in the two-site tight-binding band structure characterized by a hopping-matrix element, t , describing the bond strength between nearest-neighbour atoms separated by a distance a . Although this model is realized naturally in graphene, where carbon atoms comprise the

honeycomb structure, it is more general and can describe any atoms in the same lattice, and even electrons alone, subject to the same symmetries but devoid of atomic ‘containers’. However, tunable artificial Dirac fermions in honeycomb lattices have not been found in either atomic^{3,4} or condensed-matter^{13,14} systems.

In this work, we show that the synthesis and control of Dirac fermions in a general solid-state material is readily possible and can be interpreted in terms of simple ideas such as the decoration of crystal surfaces with atoms or molecules and periodic and aperiodic assembly, and by the application of familiar band structure models. In principle, any two-dimensional electron system (2DES) can be transformed into a host for Dirac fermions if it is patterned with a suitable periodic array of gates. In the case of top gates biased to a positive potential relative to a buried 2DES, a honeycomb arrangement of gates or dots¹⁴ is necessary to induce electrons to form a graphene structure. The dual of this configuration can also be exploited: for negative potentials that deplete electrons under the gates, a triangular gate array leaves electrons in a honeycomb pattern and is described by the same Dirac Hamiltonian. Recently we showed how single atoms can function as atomic-size gates of a 2DES at noble metal surfaces¹⁵ and how simple molecules such as CO function as repulsive potentials for surface electrons when shaped into open⁵ and closed¹⁶ quantum structures. Here we use individual CO molecules arranged molecule by molecule on a Cu(111) surface as a tunable gate array to transform a 2DES passing through these lattices into Dirac fermions, whence the term ‘molecular graphene’. The specific potentials established induce these quasi-particles to condense into various topological phases. The resulting nanostructures need not even be periodic. In fact, control over every lattice position and potential in the artificial materials we construct provides unprecedented control of the spatial texture of the hopping parameter, ultimately allowing observation of electronic ordering into ground states rarely encountered in natural systems. (See Supplementary Fig. 1 for a summary.)

The starting point of our experiment is the nearly free 2DES on Cu(111), which is characterized by very long coherence lengths ($>1,000$ Å near the Fermi energy, E_F , where the sample bias is $V = E/e = 0$ and e is the electron charge), a band edge at $E_0 = -0.45$ eV and an effective mass of $m^* = 0.38m_e$, where m_e is the electron mass. For $E < 0.5$ eV, the dispersion relation is parabolic and the Fermi velocity is $v_F = 6.45 \times 10^5$ m s⁻¹. Carbon monoxide molecules are adsorbed on clean Cu(111) and a lattice is assembled by positioning the molecules individually using the tip of a scanning tunnelling microscope (Fig. 1a and Supplementary Video 1). All experiments were performed at a temperature of 4.2 K. By applying a periodic potential at the surface, we can embed a topological singularity into the energy bands, thus forming a Dirac point and producing linear carrier dispersion. In this work, we made molecular graphene consisting of ~ 100 –1,000 molecules. For large lattices, the numbers of equivalent C sites and C–C bonds are respectively almost two and three times the number of assembled molecules.

¹Department of Physics, Stanford University, Stanford, California 94305, USA. ²Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ³Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA. ⁴Department of Applied Physics, Stanford University, Stanford, California 94305, USA. ⁵Instituto de Ciencia de Materiales de Madrid, CSIC, Cantoblanco, E-28049 Madrid, Spain.

*These authors contributed equally to this work.

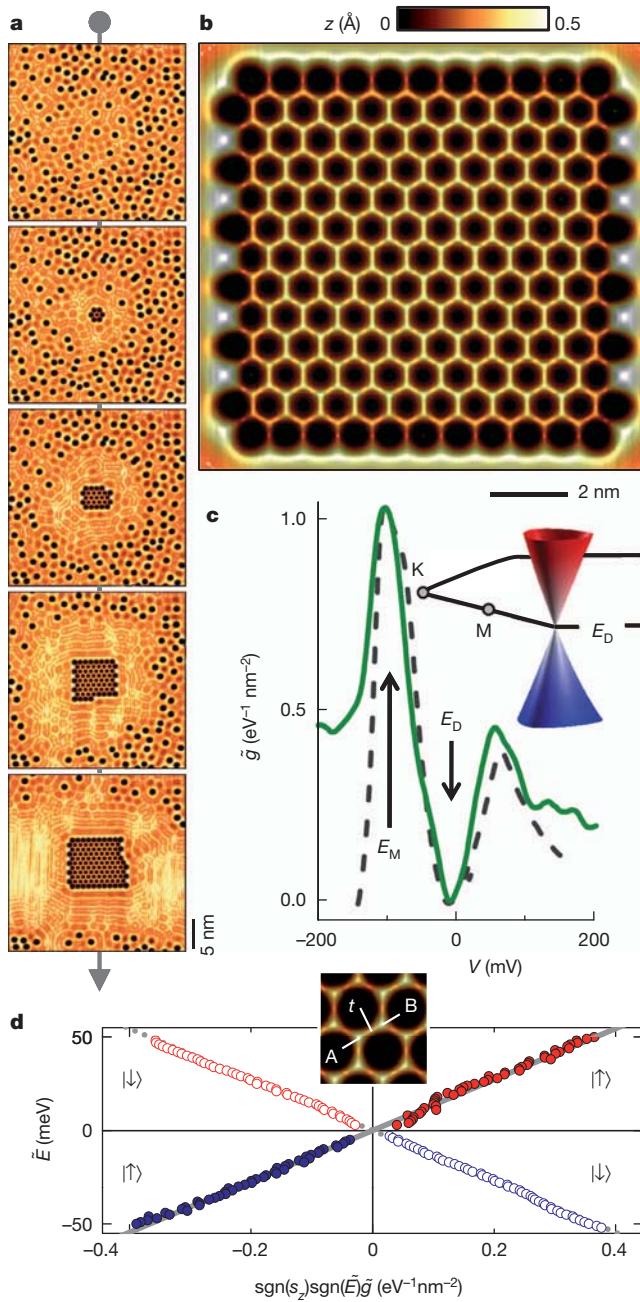


Figure 1 | Dirac fermions in molecular graphene. **a**, Sequence of constant-current topographs during the assembly of a molecular graphene lattice ($V = 10$ mV, $I = 1$ nA). **b**, Topograph of a molecular graphene lattice composed of 149 CO molecules (lattice constant, $d = 8.8$ Å). **c**, Spatially averaged, normalized differential conductance spectrum, $\tilde{g}(V)$ (solid line), measured on the top sites near the centre of quasi-neutral molecular graphene ($d = 19.2$ Å), accompanied by a tight-binding DOS fit (dashed line) with hopping parameters $t = 90$ meV and $t' = 16$ meV. Inset, resulting Dirac cone realized in reciprocal space (corresponding to fit parameters). The tight-binding spectrum is calculated by finding energy eigenvalues of a finite graphene lattice with Lorentzian basis functions (to model the finite lifetime due to scattering to bulk states and coupling to the two-dimensional continuum at the graphene edges, we used an electron self-energy $\Sigma = \Gamma/2$, where the linewidth is $\Gamma = 40$ meV from observed broadening of states near E_F). **d**, Linearly dispersing quasi-particles revealed by the conductance spectra $\tilde{g}(\tilde{E}, r)$, plotted individually for sublattice A (filled circles; pseudospin $s_z = +1/2$, $|\uparrow\rangle$) and sublattice B (open circles; pseudospin $s_z = -1/2$, $|\downarrow\rangle$), measured at locations r illustrated in the inset. Points for $|\tilde{E}| \lesssim eV_{rms}$, where V_{rms} is the modulation voltage, are excluded from this plot because this instrumental broadening prohibits their accurate measurement.

A completed ‘flake’ of molecular graphene is shown in topographic form in Fig. 1b, demonstrating a perfect internal honeycomb lattice and discernible edge effects at the termination boundaries. The spectrum shown in Fig. 1c was measured at the lattice C sites near the centre of a lattice built using 271 CO molecules separated by a distance $d = \sqrt{3}a = 19.23$ Å. The spectra in all the figures show surface-state conductance, $\tilde{g}(E, r)$, where r denotes the measurement position. (Henceforth, ‘tilde’ quantities refer to continuum properties of the Dirac fermions.) These spectra are measured by taking the ratio, g_R , between the measured differential tunnelling conductance and the spatially averaged value acquired on clean Cu(111) (Supplementary Fig. 2). This normalization removes the featureless slope present in the bare Cu spectrum and cancels the effect of possible energy-dependent tunnelling matrix elements that may vary between different microscope tips. The jump in differential conductance at the two-dimensional band edge, $g_{2D} = m^*/\pi\hbar^2 = 1.585$ eV⁻¹ nm⁻², additionally provides a quantitative calibration of the surface density of states (DOS) and is used to scale g_R to meaningful units (Supplementary Information).

The edge of the gap at the M point in momentum space (Fig. 1c) is marked by the peak in conductance at $E_M = -104$ meV. The Dirac

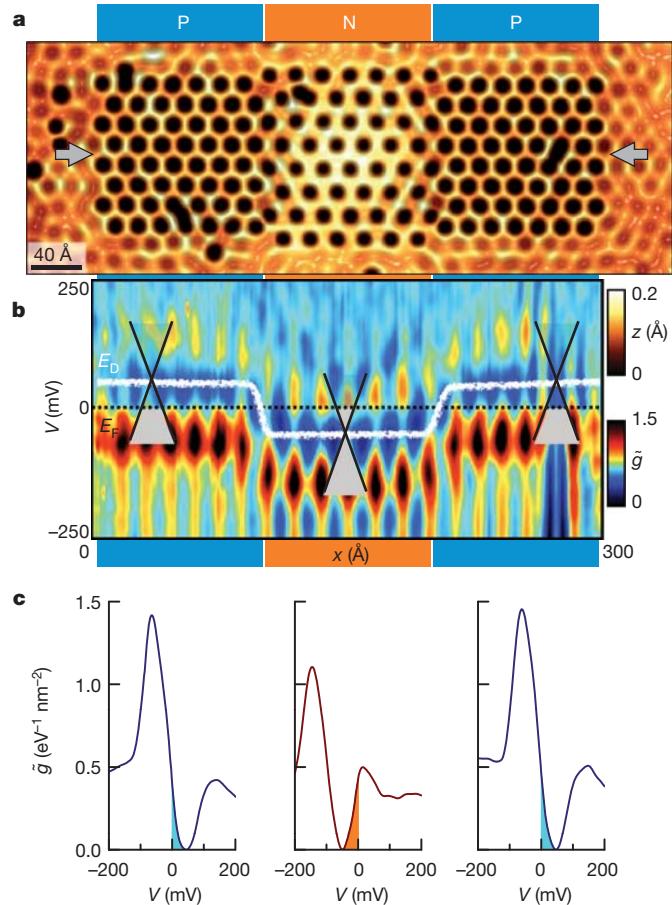
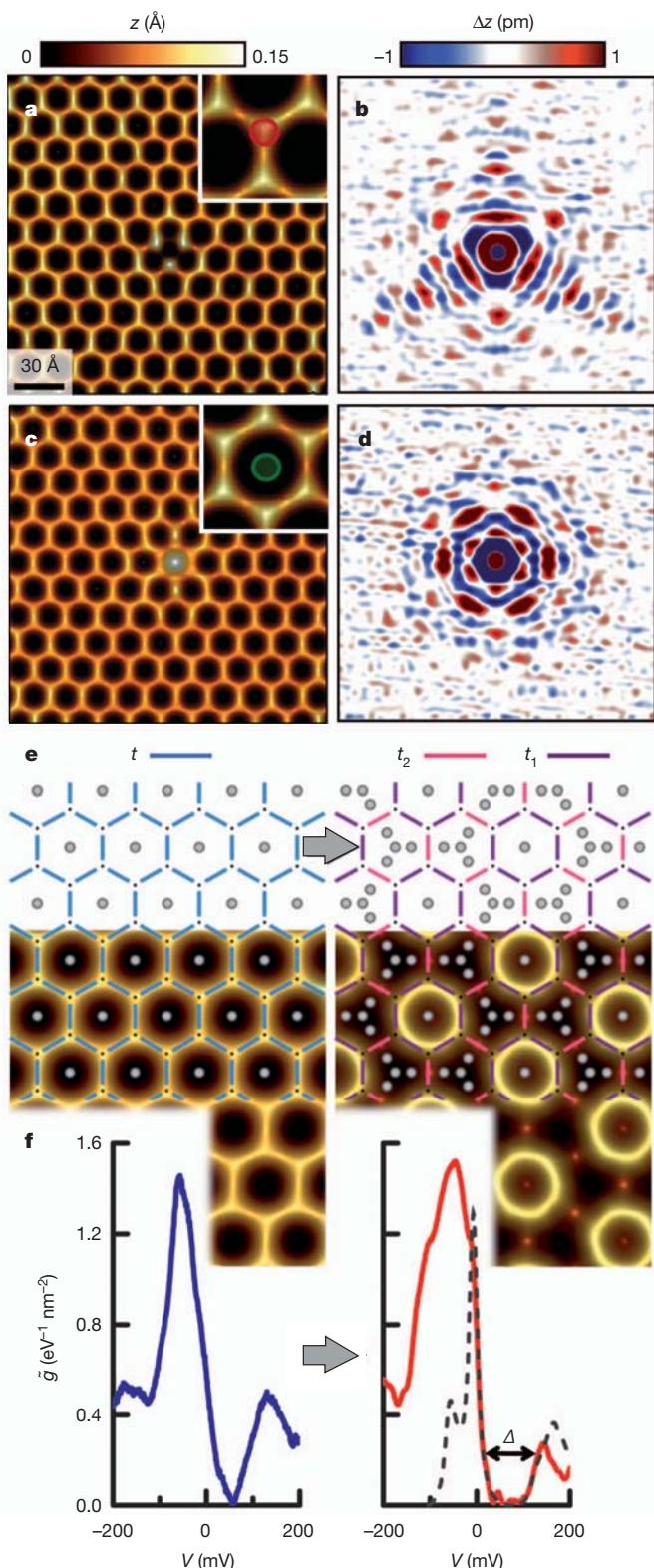


Figure 2 | Dirac point engineering in a p-n-p junction. Spectroscopic measurements made from a p-n-p lattice with alternating lattice spacings: d changes abruptly from 17.8 to 20.4 Å and then back again. **a**, Topograph of the p-n-p lattice. The conductance spectra were measured across the centre line marked by the grey arrows. **b**, Intensity colour plot of the conductance spectra $\tilde{g}(V, x)$, where x denotes the distance along the centre line. The white line is the locus of minima (the Dirac points (E_D)) in the conductance spectra. The dashed line marks the Fermi energy (E_F). Illustrative Dirac cones are superimposed to show the effective doping of each region. **c**, Spatially averaged, normalized conductance spectra measured along the centre line (marked by arrows in a). The first spectrum (blue, left) was measured in the left-hand, p-type, region ($d = 17.8$ Å), the second (orange, centre) was measured in the central, n-type, region ($d = 20.4$ Å) and the third (blue, right) was measured in the right-hand, p-type, region ($d = 17.8$ Å).

point is marked by the conductance minimum at $E_D = -5$ meV. Although the 2DES is normally decoupled from the bulk electrons for pure Cu, the presence of the molecules allows finite scattering and hence adds to the lifetime broadening evident in \tilde{g} . In a direct comparison with graphene tight binding, the nearest-neighbour hopping energy scale can be read from the data in Fig. 1c as $t \approx E_M - E_D = 99$ meV. The corresponding Dirac velocity is $\tilde{c} = 3ta/2\hbar = (2.5 \pm 0.2) \times 10^5$ m s $^{-1}$. The distinctive graphene spectral signature observed (Fig. 1c) can be contrasted



with the electronic structure of the honeycomb lattice inverse, which is a (non-Dirac) triangular electron lattice (Supplementary Fig. 3). The successful match to tight-binding theory in both cases demonstrates the promise of these techniques for realizing many other lattice types.

A complementary model uses the nearly free electron model and ‘muffin-tin’ potential¹⁷ to extract the properties of the emergent Dirac fermions. This approach (Supplementary Information) yields the position of the Dirac point, centred at the K points of the supercell Brillouin zone. For a wide range of Fourier coefficients of the superlattice potential, the nearly free electron model predicts that $t = 4\sqrt{3}\pi\hbar^2/27m^*a^2 = E_M - E_D = 118$ meV and that $\tilde{c} = 2\pi\hbar/3\sqrt{3}m^*a = (3.3 \pm 0.2) \times 10^5$ m s $^{-1}$, in good agreement with the parameters deduced from the tight-binding model.

The pseudospin structure of the Dirac point can be probed by directly tunnelling into the A and B sublattice sites¹⁷. Figure 1d shows plots of $\tilde{E} = E - E_D$ versus \tilde{g} for the two sublattices, with a sign modification depending on which sublattice is probed. This method interrogates the underlying wavefunction overlap with each pseudospin s_z , and reveals the underlying massless Dirac fermion structure in the assembled nanomaterial. The slopes of the observed linearly vanishing DOS yield $\tilde{c} = (3.1 \pm 0.2) \times 10^5$ m s $^{-1}$. This and the above values for \tilde{c} are in reasonable agreement and also match an expected theoretical value of $\tilde{c} \approx v_F/2 \approx 3.2 \times 10^5$ m s $^{-1}$ for the nearly free electron model¹⁷. Just as in real graphene¹, an electron-hole asymmetry is observed in our molecular graphene (Fig. 1c). This asymmetry is due to a finite second-nearest-neighbour hopping-matrix element, t' . A full tight-binding fit to the experiment (Fig. 1c) yields $t = 90$ meV and $t' = 16$ meV. These values agree well with the potential model (Supplementary Information). The resulting Dirac cone centred at each K point in reciprocal space is plotted in Fig. 1c (inset).

Although E_F is fixed for the underlying 2DES and cannot be changed by conventional electrostatic gating, we are able to modify the lattice parameters to change the electron count per superlattice unit cell and thus control the graphene doping level. Making these changes on the atomic scale permits a change in the Dirac fermion carrier concentration over very short distances. To demonstrate this idea, we built molecular graphene with alternating CO lattice spacings, starting with $d = 17.8$ Å and abruptly changing to 20.4 Å and then back to 17.8 Å (Fig. 2). In the lattice with the smaller spacing, the Dirac point is located 46 meV above E_F , leading to a hole-doped (p-type) lattice with a carrier density of 1.41×10^{12} cm $^{-2}$ (Fig. 2a). The lattice with the larger spacing is electron doped (n-type), with the Dirac point located 49 meV below E_F and a carrier density of 1.23×10^{12} cm $^{-2}$ (Fig. 2c). We measure the

Figure 3 | Charge- and bond-density waves in molecular graphene. a, Topographic image of quasi-neutral molecular graphene ($d = 19.2$ Å), with an additional CO molecule at the top site (location indicated in inset). b, Impurity scattering quasi-particle interference mapped through the subtraction of two topographs ($V = 10$ mV, $I = 1$ nA) measured in identical fields of view and distinguished only by the presence of the extra CO molecule at the top site. c, Topograph of quasi-neutral molecular graphene ($d = 19.2$ Å), with the central CO molecule removed from the empty site (location indicated in inset). d, Corresponding difference map of two topographs measured as above, revealing quasi-particle interference of higher symmetry. e, Left: schematic of a pure molecular graphene lattice (grey circles denote CO molecule positions, small black dots are C sites, blue lines represent uniform hopping parameter t). At bottom, the schematic is overlaid with a topograph of the corresponding experimental lattice ($d = 17.8$ Å). Right: schematic of the modifications to pure molecular graphene to obtain a Kekulé hopping texture. The addition of extra CO molecules (grey circles) splits the nearest-neighbour hopping parameter into two different values (t_1 in purple and t_2 in pink, $t_1 > t_2$), as illustrated. At bottom, the transformation schematic is overlaid with measured topography from experiment after molecular manipulation ($d = 19.2$ Å). f, Conductance spectra measured in pure molecular graphene (blue) and in the graphene lattice after the Kekulé texturing (red). The spectra display the opening of a gap, Δ , at the Dirac point. Dashed curve, tight-binding fit calculated with $t_1 = 2t_2$. The small difference in d compensates for the change in chemical potential introduced by the extra CO molecules.

surface spectrum (\tilde{g}) across the lattice along the line indicated (Fig. 2a, b, arrows), crossing all regions. Because there are no charging effects, the interface between the p- and n-type regions is very narrow: it is about 20 Å wide (Fig. 2b). The extremely short transition between the p- and n-type regions makes this device a suitable candidate in which to study phenomena such as the Klein paradox¹⁸ or to create a Veselago lens¹⁹.

Using atomic manipulation, we created both pseudospin-conserving and pseudospin-breaking local disturbances further to reveal the Dirac nature of our system. We started with quasi-neutral molecular graphene with E_D near E_F (271 CO molecules, $d = 19.2$ Å) and tested two symmetries (a C-site defect (Fig. 3a), which locally imbalances the sublattices, and an empty-site vacancy (Fig. 3c), which locally preserves sublattice symmetry); these two structures are topologically distinct because their potentials correspond respectively to a local vector potential and a local scalar potential coupling to pseudospin. By subtracting two low-bias scanning tunnelling microscope topographs, one with the impurity and the other without it, and both locked to exactly the same area and using identical measurement parameters, we obtain detailed pseudospin maps resulting from the tiny DOS perturbations caused by quantum interference. The distinct patterns observed have the three-fold symmetry (Fig. 3b) predicted for single impurities in graphene that disrupt the Berry phase²⁰, rather than the full six-fold symmetry of a scalar perturbation, which conserves pseudospin^{21–23} (Fig. 3d).

Topological changes fundamentally alter the lattice symmetry and are the key to unlocking physical phenomena such as electron fractionalization^{6–8,10–12,22}. In graphene, one of the simplest (yet unrealized) deformations is the Kekulé distortion rooted in the historical interpretation of benzene. This distortion breaks the bond symmetry of graphene by forming two hopping elements, t_1 and t_2 , in the pattern shown in Fig. 3e. We produce this distortion using a special ‘Mercedes’ arrangement of CO molecules in the honeycomb empty sites. This has the effect of modulating the strength of every other C–C bond along the perimeter of each cell. Such a distortion adds an off-diagonal term to the Hamiltonian¹¹, such that

$$H_G = \begin{pmatrix} \hbar\tilde{c}\sigma\cdot\mathbf{k} & \Delta I_{2 \times 2} \\ \Delta^* I_{2 \times 2} & -\hbar\tilde{c}\sigma\cdot\mathbf{k} \end{pmatrix}$$

where $I_{2 \times 2}$ is the two-by-two identity matrix and an asterisk denotes complex conjugate. This distortion is predicted to open an energy gap, Δ , even if the underlying sublattice symmetry is not broken; notably, this effect has never been observed. Figure 3f proves that the Kekulé distortion works as theoretically predicted, creating massive Dirac fermions out of the massless Dirac fermions in the pristine lattice. From fits to theory, the mass of the emergent fermions is $m_D = 0.1 \pm 0.02 m_e$. The Kekulé ground state^{10–12,24,25} has an intriguing mapping to a superconducting topological surface state²⁶, after pseudospin is mapped to spin and the valley degree of freedom to an isospin, equivalent to attaching a scalar gauge field that produces a Dirac fermion mass. This scalar gauge field is manifest in the bond-density wave mosaic structure visible in Fig. 3e. The transition from massless to massive Dirac fermions has been theoretically cast as a quantum phase transition^{3,11}; the molecular graphene system provides an experimental test bed of these ideas starting with the spontaneous generation of mass observed here.

The chiral character of the electronic charge in graphene is due to the pseudospin associated with the symmetry between the two triangular sublattices that form the honeycomb lattice. It has been proposed that by breaking this sublattice symmetry through strain, it is possible to generate a pseudomagnetic field and therefore obtain Landau levels and quantum Hall phases without breaking time reversal symmetry. The effect of strain has recently been observed in graphene nanobubbles²⁷, but tunable molecular graphene offers much more precise and *in situ* control over internal gauge fields. The strain field displacements in polar coordinates (r and θ) suggested⁹ to generate a constant field are $(u_r, u_\theta) = (qr^2 \sin(3\theta), qr^2 \cos(3\theta))$, where q is a parameter denoting the strength of the strain. In our final experiment, we applied this strain field to molecular graphene by means of atomic manipulation.

Topographs for successively strained graphene are shown in Fig. 4a. The value of the pseudomagnetic field can be estimated as $\tilde{B} = 8\beta\hbar q/ea = 16\pi\hbar q/3de$, where $\beta = -\partial\ln(t)/\partial\ln(a) \approx 2$ (Supplementary Information). We study strain values up to $q = 10^{-3}$ Å⁻¹, which is

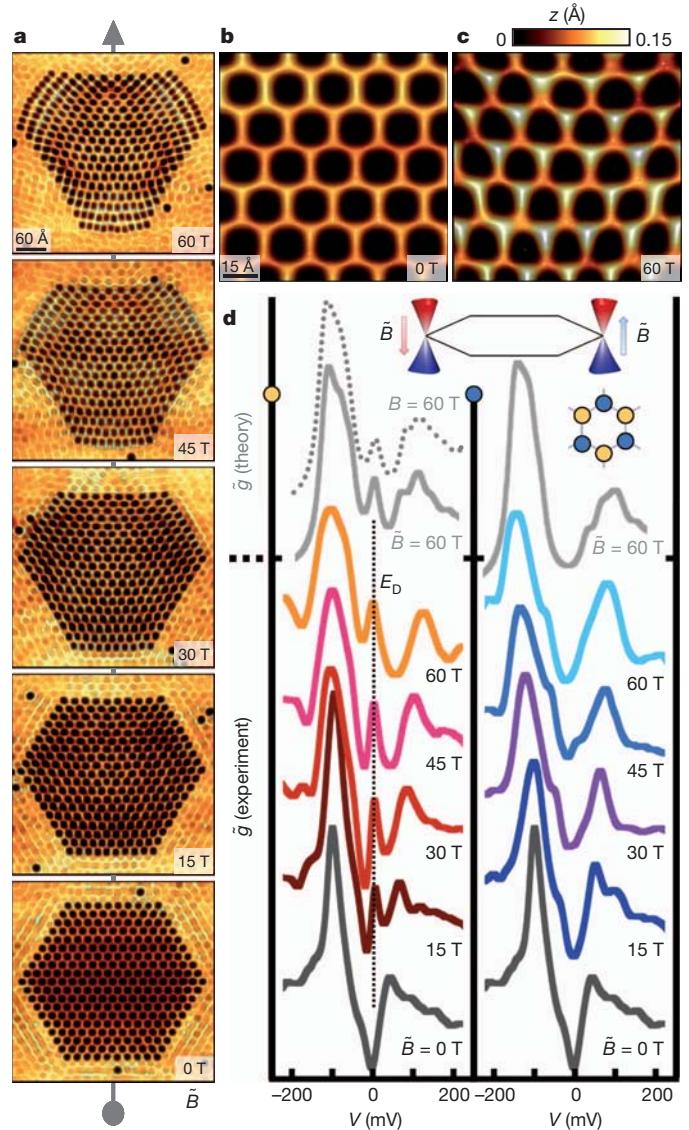


Figure 4 | Landau quantization and topological zero modes in a tunable pseudomagnetic field. **a**, Sequence of topographs of molecular graphene lattices with increasing values of triaxial strain. The position of each CO molecule was determined by the dislocation vector defined in the main text. From bottom to top, $q = 0, 2.5 \times 10^{-4}, 5 \times 10^{-4}, 7.5 \times 10^{-4}$ and 10^{-3} Å⁻¹. The corresponding values of the constant pseudomagnetic field are $\tilde{B} = 0, 15, 30, 45$ and 60 T (felt in opposite directions by the two graphene valleys; see **d**, top inset). **b**, Topograph at the centre of the lattice without strain distortion ($q = 0$ Å⁻¹), showing the unbroken symmetry between each sublattice (pseudospin) of the honeycomb. **c**, Topograph at the centre of the lattice with strain distortion ($q = 10^{-3}$ Å⁻¹), showing the broken symmetry between each sublattice (one bright and one dark) as a result of the localization of the zero Landau level on half of the sample (bright sublattice). **d**, Left: normalized conductance spectra measured on sublattice A (brighter top sites in **c** and orange circles in inset schematic) for successive values of strain. The spectra were measured near the centres of the lattices shown in **a**. Right: spectra measured on sublattice B (darker top sites in **c** and blue circles in inset schematic) for the same successive values of strain, showing the opening of a Landau gap. Grey solid curves are tight-binding fits of strained finite lattices to the experiment. The grey dotted curve shows a tight-binding calculation in a real magnetic field, $B = 60$ T, for an unstrained lattice of the same dimensions and hopping parameters.

equivalent to a pseudomagnetic field of $\tilde{B} \approx 60$ T, corresponding now to a vector gauge field dressing the Dirac fermions (the applied triaxial strain can be contrasted with uniaxial strain (Supplementary Fig. 4), which is devoid of gauge fields). The breaking of the pseudospin symmetry of the Dirac point can be probed by direct tunnelling into the individual sublattice sites. Once the strain is applied, the symmetry breaking between the two sublattices is discernible in the topographs (Fig. 4b, c) and can be noticed in the spectra (Fig. 4d). In the 'bright' sublattice (A), a well-defined zero-bias state—the zero Landau level—forms and gains prominence with increasing strain. In the 'dark' sublattice (B), the spectra are characterized by a reduction in the conductance at zero bias with strain—the formation of a Landau gap—visible through the transformation of typical V-shaped spectra into U-shaped contours.

The presence of a zeroth Landau level at the Dirac point is one of the hallmarks of the quantum Hall effect with Dirac fermions. The low Fermi velocity here means that even a small strain can push the first excited Landau level beyond the linear regime in the band structure, towards or beyond the edge of the M-point gap. We estimate that for our lowest value of strain ($q = 0.25 \times 10^{-3} \text{ \AA}^{-1}$), $\tilde{B} \approx 15$ T and the $n = 1$ Landau level therefore lies at 33 mV. The registry of the CO molecules to the surface Cu atoms quantizes the allowed lattice sizes and hinders the formation of lattices with smaller strains. Nevertheless, a complete quantitative model of our finite system shows that it behaves as a graphene quantum dot in a real magnetic field, B , when these strain fields are applied (in Fig. 4d, tight-binding fits for finite strain and zero B and for zero strain and high B both provide excellent reproductions of the observations).

Owing to the large strains applied, the particular topological state observed here is an ultraquantized ground state: only the zeroth Landau level is present and fractionally occupied by Dirac fermions. Such a state is unprecedented in normal graphene owing to the larger energy scales and, hence, very large magnetic fields necessary to study it. The pseudospin symmetry breaking observed is connected with the fundamental index theorem, which predicts the same topological phase for Dirac fermions on a lattice²⁸ as is observed here. In this phase, the electrons in the zero mode form a peculiar condensate in which they become completely delocalized over the sample but remain alternately localized and antilocalized in the pseudospin projection²⁹ (Supplementary Information). The image in Fig. 4c provides a snapshot of this intriguing phase in real space.

The study of artificial lattices may lead to technological applications, but they also provide a new level of control over Dirac fermions and allow experimental access to a set of phenomena that has hitherto been considered only theoretically. The introduction of tunable interactions between the electrons could lead to the formation of spin liquids in graphene, and the addition of spin-orbit coupling could lead to a quantum spin Hall effect. The time-reversal-invariant quantum Hall state observed here already has a direct connection to the two-dimensional topological insulator⁹. In future, such topological phases may be produced even more easily by extending our manipulation techniques to self-assembly methods³⁰, which naturally generate long-range periodic potentials on decorated surfaces.

Received 8 July 2011; accepted 9 February 2012.

- Castro Neto, A., Guinea, F., Peres, N., Novoselov, K. & Geim, A. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109–162 (2009).
- Bloch, I. Ultracold quantum gases in optical lattices. *Nature Phys.* **1**, 23–30 (2005).
- Zhu, S.-L., Wang, B. & Duan, L.-M. Simulation and detection of Dirac fermions with cold atoms in an optical lattice. *Phys. Rev. Lett.* **98**, 260402 (2007).

- Wunsch, B., Guinea, F. & Sols, F. Dirac-point engineering and topological phase transitions in honeycomb optical lattices. *N. J. Phys.* **10**, 103027 (2008).
- Moon, C. R., Mattos, L. S., Foster, B. K., Zeltzer, G. & Manoharan, H. C. Quantum holographic encoding in a two-dimensional electron gas. *Nature Nanotechnol.* **4**, 167–172 (2009).
- Chamon, C. et al. Irrational versus rational charge and statistics in two-dimensional quantum systems. *Phys. Rev. Lett.* **100**, 110405 (2008).
- Seradjeh, B., Weeks, C. & Franz, M. Fractionalization in a square-lattice model with time-reversal symmetry. *Phys. Rev. B* **77**, 033104 (2008).
- Ryu, S., Mudry, C., Hou, C.-Y. & Chamon, C. Masses in graphenelike two-dimensional electronic systems: topological defects in order parameters and their fractional exchange statistics. *Phys. Rev. B* **80**, 205319 (2009).
- Guinea, F., Katsnelson, M. I. & Geim, A. K. Energy gaps and a zero-field quantum Hall effect in graphene by strain engineering. *Nature Phys.* **6**, 30–33 (2010).
- Jackiw, R. Fractional charge from topology in polyacetylene and graphene. *AIP Conf. Proc.* **939**, 341–350 (2007).
- Hou, C.-Y., Chamon, C. & Mudry, C. Electron fractionalization in two-dimensional graphenelike structures. *Phys. Rev. Lett.* **98**, 186809 (2007).
- Seradjeh, B. & Franz, M. Fractional statistics of topological defects in graphene and related structures. *Phys. Rev. Lett.* **101**, 146401 (2008).
- Cahangirov, S., Topsakal, M., Aktürk, E., Şahin, H. & Ciraci, S. Two- and one-dimensional honeycomb structures of silicon and germanium. *Phys. Rev. Lett.* **102**, 236804 (2009).
- Singha, A. et al. Two-dimensional Mott-Hubbard electrons in an artificial honeycomb lattice. *Science* **332**, 1176–1179 (2011).
- Moon, C. R., Lutz, C. P. & Manoharan, H. C. Single-atom gating of quantum-state superpositions. *Nature Phys.* **4**, 454–458 (2008).
- Moon, C. R. et al. Quantum phase extraction in isospectral electronic nanostructures. *Science* **319**, 782–787 (2008).
- Park, C.-H. & Louie, S. G. Making massless Dirac fermions from a patterned two-dimensional electron gas. *Nano Lett.* **9**, 1793–1797 (2009).
- Katsnelson, M., Novoselov, K. & Geim, A. Chiral tunnelling and the Klein paradox in graphene. *Nature Phys.* **2**, 620–625 (2006).
- Cheianov, V., Fal'ko, V. & Altshuler, B. The focusing of electron flow and a Veselago lens in graphene p-n junctions. *Science* **315**, 1252–1255 (2007).
- Wehling, T. O. et al. Theory of Fano resonances in graphene: the influence of orbital and structural symmetries on STM spectra. *Phys. Rev. B* **81**, 085413 (2010).
- Wehling, T. O. et al. Local electronic signatures of impurity states in graphene. *Phys. Rev. B* **75**, 125425 (2007).
- Guinea, F., Katsnelson, M. & Vozmediano, M. Midgap states and charge inhomogeneities in corrugated graphene. *Phys. Rev. B* **77**, 075422 (2008).
- Bena, C. & Kivelson, S. A. Quasiparticle scattering and local density of states in graphite. *Phys. Rev. B* **72**, 125432 (2005).
- Martin, I., Blanter, Y. & Morpurgo, A. Topological confinement in bilayer graphene. *Phys. Rev. Lett.* **100**, 036804 (2008).
- Nomura, K., Koshino, M. & Ryu, S. Topological delocalization of two-dimensional massless Dirac fermions. *Phys. Rev. Lett.* **99**, 146806 (2007).
- Roy, B. & Herbut, I. Unconventional superconductivity on honeycomb lattice: theory of Kekulé order parameter. *Phys. Rev. B* **82**, 035429 (2010).
- Levy, N. et al. Strain-induced pseudo-magnetic fields greater than 300 Tesla in graphene nanobubbles. *Science* **329**, 544–547 (2010).
- Kailasvuori, J. Pedestrian index theorem à la Aharonov-Casher for bulk threshold modes in corrugated multilayer graphene. *Europhys. Lett.* **87**, 47008 (2009).
- Herbut, I. Pseudomagnetic catalysis of the time-reversal symmetry breaking in graphene. *Phys. Rev. B* **78**, 205433 (2008).
- Lobo-Checa, J. et al. Band formation from coupled quantum dots formed by a nanoporous network on a copper surface. *Science* **325**, 300–303 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under contract DE-AC02-76SF00515. F.G. acknowledges financial support from MICINN (Spain) through grants FIS2008-00124 and CONSOLIDER CSD2007-00010, and calculations supported by the US National Science Foundation. We thank C.-H. Park, I. Martin, A. Balatsky, T. Wehling, A. Akhmerov, E. Heller and A. Fetter for discussions.

Author Contributions K.K.G., W.M. and W.K. designed and performed experiments, analysed data and wrote the manuscript. F.G. provided the theoretical analysis. H.C.M. directed the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.C.M. (manoharan@stanford.edu).

DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response

Anton Kuzyk^{1*}†, Robert Schreiber^{2*}, Zhiyuan Fan³, Günther Pardatscher¹, Eva-Maria Roller², Alexander Högele², Friedrich C. Simmel¹, Alexander O. Govorov³ & Tim Liedl²

Matter structured on a length scale comparable to or smaller than the wavelength of light can exhibit unusual optical properties¹. Particularly promising components for such materials are metal nanostructures, where structural alterations provide a straightforward means of tailoring their surface plasmon resonances and hence their interaction with light^{2,3}. But the top-down fabrication of plasmonic materials with controlled optical responses in the visible spectral range remains challenging, because lithographic methods are limited in resolution and in their ability to generate genuinely three-dimensional architectures^{4,5}. Molecular self-assembly^{6,7} provides an alternative bottom-up fabrication route not restricted by these limitations, and DNA- and peptide-directed assembly have proved to be viable methods for the controlled arrangement of metal nanoparticles in complex and also chiral geometries^{8–14}. Here we show that DNA origami^{15,16} enables the high-yield production of plasmonic structures that contain nanoparticles arranged in nanometre-scale helices. We find, in agreement with theoretical predictions¹⁷, that the structures in solution exhibit defined circular dichroism and optical rotatory dispersion effects at visible wavelengths that originate from the collective plasmon–plasmon interactions of the nanoparticles positioned with an accuracy better than two nanometres. Circular dichroism effects in the visible part of the spectrum have been achieved by exploiting the chiral morphology of organic molecules and the plasmonic properties of nanoparticles^{18–20}, or even without precise control over the spatial configuration of the nanoparticles^{12,21,22}. In contrast, the optical response of our nanoparticle assemblies is rationally designed and tunable in handedness, colour and intensity—in accordance with our theoretical model.

To demonstrate the potential of DNA origami for the programmable and nanometre-precise design of plasmonic nanostructures, we targeted an optical response that emerges as a collective effect from spatially precisely arranged multiple nanoparticles in close proximity, and that requires genuinely three-dimensional structures. The response of helically arranged metallic nanocrystals to an electromagnetic field meets both criteria. Much like the circular dichroism (CD, the differential absorption of left and right circularly polarized light) effect seen with ‘optically active’ chiral molecules such as DNA and proteins that exhibit CD in the ultraviolet and infrared ranges owing to the electronic and vibronic excitations of their chiral secondary structure²³, strong CD signals in the visible spectrum have been predicted to occur through collective Coulomb interaction of plasmonic dipoles in chiral assemblies of metal nanoparticles¹⁷. (See Supplementary Information notes 1 and 2 for information on naturally occurring CD, and on the theoretical framework of plasmon-induced CD.)

To create helical nanoparticle assemblies, we used the DNA origami^{15,16} approach where hundreds of rationally designed ‘staple’ oligonucleotides are hybridized to a long single-stranded DNA scaffold

to force it into a specific two- or three-dimensional shape. The resulting objects are fully addressable by their DNA sequence, enabling further decoration and functionalization in a unique, sequence-specific manner²⁴. Geometrical and material parameters can be tailored by simple adjustments of the design and fabrication protocols, which should make it feasible to quantitatively tune the optical activity of such nanostructures dispersed in solution.

Figure 1a illustrates the nanostructure design (design details and DNA sequences are given in Supplementary Information note 3). It is based on DNA origami 24-helix bundles that offer nine helically arranged attachment sites for plasmonic particles coated with single-stranded DNA, here gold nanoparticles with a diameter of 10 nm. By using nanoparticles covered with multiple linker strands, we can avoid yield-lowering purification procedures and instead add an excess of modified gold particles for hybridization to the DNA structures; unconjugated gold particles are simply removed for re-use in later experiments. The quality of the assemblies was assessed by transmission electron microscopy (TEM) (Fig. 1b) (experimental details and additional TEM images can be found in Supplementary Information note 4). The strength of the optical activity critically depends on the quality of the assembled structures²⁵, so we improved our experimental protocol to achieve a nanoparticle attachment yield of 96% and 98% per site and an overall yield of 77% and 86% of perfect assemblies of right- and left-handed helices with nine nanoparticles, respectively (Supplementary Information note 5). Most of the imperfect structures exhibited only a single defect within an otherwise well-formed helix. Helices stored for three months at 4 °C were still intact (Supplementary Information note 6). As illustrated in Fig. 1c, CD measurements were then carried out on solutions containing left- or right-handed nanohelices.

Left-handed helical arrangements of plasmonic particles are expected to produce a CD signal with a characteristic bisignate peak–dip shape, with the peak centred around the plasmon resonance frequency. Right-handed helices should produce a vertically mirrored dip–peak signal. Indeed, CD spectra measured on samples with gold nanoparticle assemblies of both helicities exhibited the anticipated signatures (Fig. 2a). These responses arise because the helical arrangement of the gold nanoparticles results in coupled plasmon waves propagating along a helical path and causing increased absorption of those components of the incident light that are in accord with the handedness of the helices. The bisignate appearance of the CD signals results from the isotropic nature of our samples, in which the nanohelices are randomly oriented with respect to the direction of the incident light. The interactions between the individual gold particles within each helix create a splitting between the longitudinal and transverse modes of the electromagnetic wave, and these modes typically have opposite chirality. As a result, the plasmonic CD spectrum acquires the characteristic dip–peak shape (Supplementary Information note 2).

¹Physik Department and ZNN/WSI, Technische Universität München, Am Coulombwall 4a, 85748 Garching, Germany. ²Fakultät für Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany. ³Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701, USA. †Present address: Department of Applied Physics, Aalto University School of Science, FI-00076 Aalto, Finland.

*These authors contributed equally to this work.

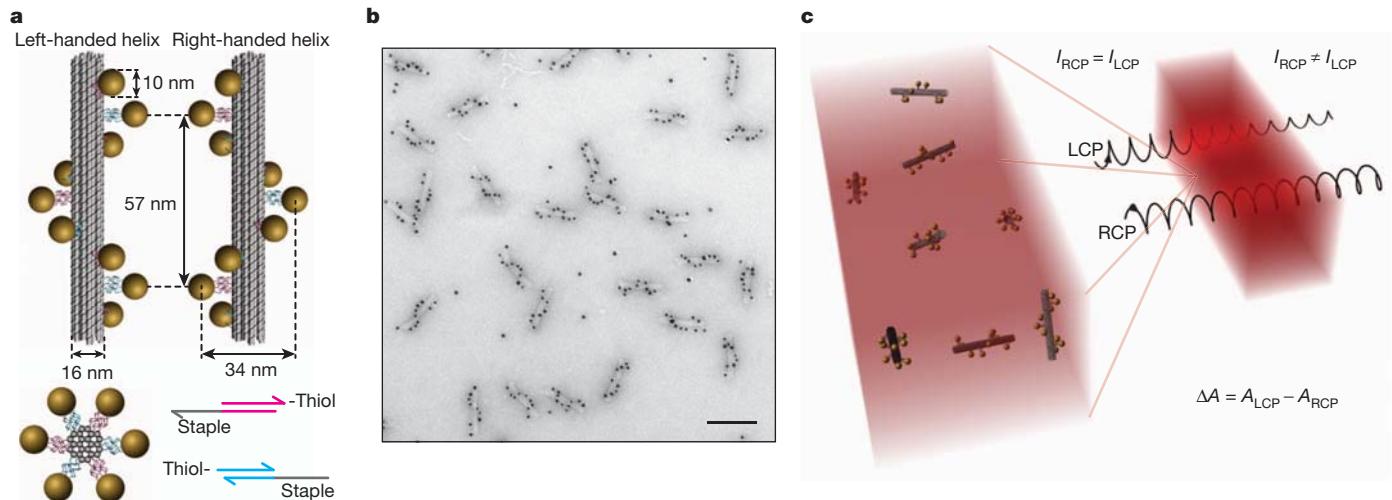


Figure 1 | Assembly of DNA origami gold nanoparticle helices and principle of circular dichroism. **a**, Left- and right-handed nanohelices (diameter 34 nm, helical pitch 57 nm) are formed by nine gold nanoparticles each of diameter 10 nm that are attached to the surface of DNA origami 24-helix bundles (each of diameter 16 nm). Each attachment site consists of three 15-nucleotide-long single-stranded extensions of staple oligonucleotides. Gold nanoparticles carry multiple thiol-modified DNA strands, which are complementary to these staple extensions. Nanoparticles and 24-helix bundles

are mixed for assembly and the resulting constructs are gel-purified. **b**, TEM image of assembled left-handed gold nanohelices (scale bar, 100 nm). Analysis of the TEM data yields a 98% success rate for directed attachment of nanoparticles. **c**, Circular dichroism is measured as the difference in absorbance $\Delta A = A_{LCP} - A_{RCP}$ of left-hand-circularly polarized (LCP) and right-hand-circularly polarized (RCP) light as a function of wavelength. CD measurements were performed with a CD spectrometer on samples in cuvettes of optical path length 3 mm.

The experimental spectra agree well with theoretical calculations based on classical electrodynamics (Fig. 2b). The strongest absorption of circularly polarized light is predicted in the vicinity of the surface plasmon frequency of the metal nanoparticles, as experimentally seen with our assemblies. Strikingly, even the strength of the measured signals closely matches the magnitude predicted by the calculations. We further confirmed theoretically that the CD signal remains robust to positional perturbations of the individual particles (see Supplementary Information note 2).

The dipole theory¹⁷ predicts that the CD signal becomes stronger when the particles are either larger or arranged in a tighter helix:

$$CD_{\text{plasmon}} \approx \frac{a_{\text{NP}}^{12}}{R_0^8} \quad (1)$$

where a_{NP} and R_0 are the nanoparticle and helix radii, respectively. We quantified the influence of particle size on the CD signal by using electroless deposition of metal from solution^{26,27} onto 10-nm seeding particles that were already assembled into the helical geometry described in Fig. 1 (see Supplementary Information note 7 for experimental details and additional TEM images). CD measurements of these ‘enhanced’ samples showed two notable effects (Fig. 2c): the signal strength increased up to 400-fold for nanoparticle diameters of 16 ± 2 nm, and the absorption as well as the CD peak shifted to longer wavelengths. These results are consistent with theoretical predictions for the plasmonic CD effect. A simple estimate for the CD signal increase based on equation (1) and assuming a particle size of 16 nm yields an enhancement of about 280, and quantitative numerical calculations predict an enhancement of about 500. As in the experiments, the calculated CD for larger particles and the same helix radius is also red-shifted (Fig. 2d), which is typical for strongly interacting plasmonic nanocrystals²⁸. Taken together, these observations illustrate that desired optical behaviour can be designed and realized by exploiting the collective interactions between plasmonic particles attached with close to 100% yield and positioned with nanometre-scale precision.

To generate CD at other wavelengths, we plated silver onto pre-assembled 10-nm gold particle helices and thereby produced a silver shell of about 3 nm around each of the gold nanoparticles. Plasmon resonances in silver occur at a shorter wavelength than in gold, so the

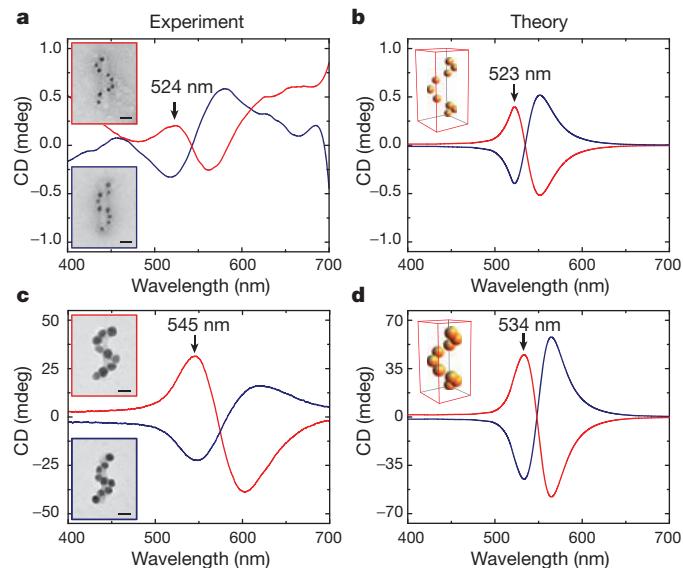


Figure 2 | Circular dichroism of self-assembled gold nanohelices. Experimental (a and c) and theoretical (b and d) CD spectra of left-handed (red lines) and right-handed (blue lines) helices of nine gold nanoparticles show characteristic bisignate signatures in the visible. **a**, CD spectra of nanohelices composed of 10-nm gold particles. The peak position in the spectra of left-handed helices (indicated by the arrow) coincides with the plasmon absorption resonance maximum. **b**, The theoretically predicted CDs for the geometries in a exhibit the same features; the peak positions and amplitudes are in remarkable agreement with the experiment. **c**, The CD signal increases owing to collective plasmonic enhancement by a factor of 400 for assemblies of nanoparticles with 16-nm diameter, rendering the noise in the spectra invisible (as in a). The peak position for left-handed helices exhibits a red-shift from 524 nm to 545 nm. **d**, The corresponding theoretical calculation predicts a 500-fold enhancement of the signal and a peak shift from 523 nm to 534 nm. The CD spectra were recorded at concentrations of nanohelices of 1.5 nM in a and 0.4 nM in c. The insets in a and c show TEM images of left-handed (red frame) and right-handed (blue frame) nanohelices (scale bars, 20 nm); the respective left-handed model geometries are depicted in the insets to b and d.

recorded CD spectra were shifted into the blue spectral region (Fig. 3a). To fine-tune the optical response to intermediate wavelengths, we used mixtures of gold and silver ions in the plating solution to grow alloy shells around the seed particles. Figure 3a presents a TEM image (upper inset) with the corresponding CD spectrum of an alloy-coated structure. As expected, the CD signal for the alloy helices appears at an intermediate wavelength (absorption spectra and additional CD spectra are presented in Supplementary Information note 8). Mixing of solutions containing different types of structures resulted in a linear superposition of the corresponding spectra (Fig. 3a). These experiments indicate the potential of our assembly method for the fabrication of optically active materials with customized spectral response.

To visualize the collective optical activity of our self-assembled nanohelices in a macroscopic optical experiment, we exposed gold nanoparticle helices to excessive amounts of silver ions during electroless metal deposition and thereby achieved a giant molar CD of around $10^8 \text{ M}^{-1} \text{ cm}^{-1}$. The corresponding spectra are presented in Fig. 3b. Droplets of samples containing left-handed or right-handed helices and a non-helical control droplet were placed onto a glass slide between two linear polarizers. As demonstrated in Fig. 4 with polarization-resolved transmission images, the droplets with gold helices rotate the polarization angle of the linearly polarized light. Although our set-up is insensitive to polarization below 550 nm, rotation of the polarization axis of the linearly polarized light was detected for both right- and left-handed helices in the red band of the visible spectrum. The observation is in accordance with the effect of optical rotatory dispersion, which is characteristic of chiral materials exhibiting CD. In the case of our right-handed "metafluid"²⁹, the polarization axis is rotated clockwise (or counter-clockwise) at wavelengths above (or below) 560 nm; the opposite effect is seen with the left-handed nanohelices. The control sample containing isotropically dispersed 10-nm gold nanoparticles exhibited no such optical activity. Further details on the materials characteristics (anisotropy factor and specific rotation) can be found in Supplementary Information notes 8 and 9.

In future experiments we will explore the possibility of realizing materials with a negative refractive index without the requirement of

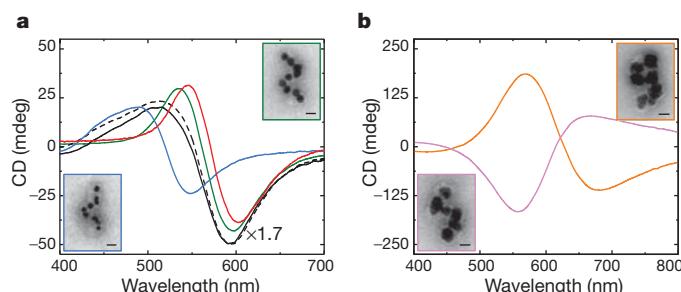


Figure 3 | Spectral tuning of circular dichroism by metal composition. **a**, A pronounced blueshift of the peak position in the CD response was achieved by depositing several nanometres of silver on gold nanoparticles of diameter 10 nm (blue spectrum of left-handed gold-core/silver-shell nanohelices; red spectrum from Fig. 2c is shown for reference). Growth of silver+gold alloy shells gives rise to an intermediate blueshift (green spectrum of left-handed gold-core/silver+gold-shell helices). Mixing of solutions that contain nanohelices of various metal compositions results in a superposition of individual CD responses (black spectrum): a 2:3 mixture of gold-core/silver-shell (blue spectrum) and gold-core/silver+gold-shell (green spectrum) exhibits CD close to the predicted superposition (dashed black line; black curves were rescaled by a factor of 1.7 for presentation purposes). **b**, Excessive metal deposition leads to the merging of nanoparticles and further enhancement of the CD signal for the left-handed (orange spectrum) and right-handed (purple spectrum) gold-core/silver-shell helices. The corresponding molar CD in the visible spectrum is huge with a strength of about $10^8 \text{ M}^{-1} \text{ cm}^{-1}$. The CD spectra were recorded at concentrations of nanohelices of 0.4 nM in **a** and 40 pM in **b**. The insets in **a** and **b** show TEM images of left-handed (blue, green and orange frames) and right-handed (purple frame) nanohelices (scale bars, 20 nm).

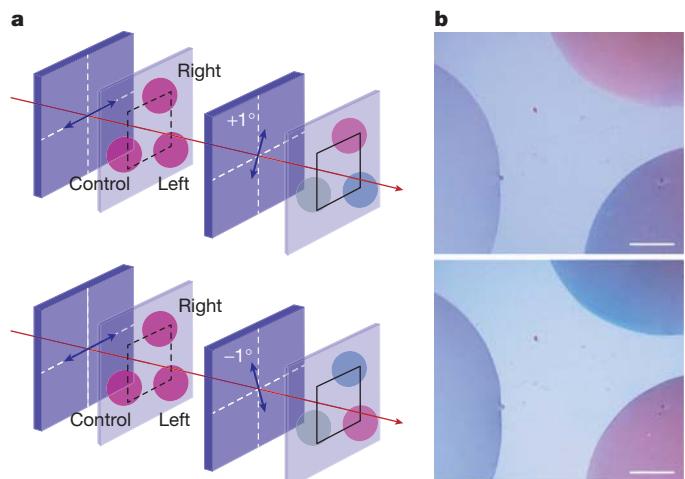


Figure 4 | Optical rotatory dispersion of self-assembled gold nanohelices.

a, The macroscopic optical activity of the metafluid is demonstrated in polarization-resolved transmission by placing droplets of left-handed and right-handed nanohelices as well as a control droplet with isotropically dispersed gold nanoparticles between two linear polarizers. The components of the transmission set-up are ordered along the propagation path of white light as follows: linear polarizer, sample, linear analyser, screen (charge-coupled-device camera). The extinction characteristics of the linear polarizers render the set-up polarization-insensitive for wavelengths below 550 nm. **b**, Real-colour transmission images (scale bar, 1 mm) of the droplets for clockwise (top) and counter-clockwise (bottom) rotation of the analyser by one degree out of the orthogonal configuration. The droplets of right-handed and left-handed helices appear red in the top and bottom images, respectively, as a result of optical rotatory dispersion. Right- (or left-) handed nanohelices rotate the polarization axis of the transmitted light in positive (or negative) and negative (or positive) directions in the red and blue spectral bands, respectively. The rotation of the polarization in the red band is sensitively detected by the set-up, but green and blue components of white light illumination (below 550 nm) are not polarization-resolved and contribute only to the bluish background.

either ϵ or μ being negative³⁰, using fluids with oriented or densely packed plasmonic helices, which should lead to a further increase of the chiral optical response. But irrespective of whether this goal can be realized, the present findings clearly establish DNA origami as a valuable addition to the existing tools at the material engineer's disposal for precisely arranging nanoparticles into assemblies with desired electric or magnetic properties.

Received 12 September 2011; accepted 17 January 2012.

1. Liu, Y. & Zhang, X. Metamaterials: a new frontier of science and technology. *Chem. Soc. Rev.* **40**, 2494–2507 (2011).
2. Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424**, 824–830 (2003).
3. Polman, A. Plasmonics applied. *Science* **322**, 868–869 (2008).
4. Soukoulis, C. M. & Wegener, M. Past achievements and future challenges in the development of three-dimensional photonic metamaterials. *Nature Photon.* **5**, 523–530 (2011).
5. Gansel, J. K. *et al.* Gold helix photonic metamaterial as broadband circular polarizer. *Science* **325**, 1513–1515 (2009).
6. Jones, M. R., Osberg, K. D., Macfarlane, R. J., Langille, M. R. & Mirkin, C. A. Templated techniques for the synthesis and assembly of plasmonic nanostructures. *Chem. Rev.* **111**, 3736–3827 (2011).
7. Fan, J. A. *et al.* Self-assembled plasmonic nanoparticle clusters. *Science* **328**, 1135–1138 (2010).
8. Seeman, N. C. Nanomaterials based on DNA. *Annu. Rev. Biochem.* **79**, 65–87 (2010).
9. Tan, S. J., Campolongo, M. J., Luo, D. & Cheng, W. Building plasmonic nanostructures with DNA. *Nature Nanotechnol.* **6**, 268–276 (2011).
10. Ding, B. *et al.* Gold nanoparticle self-similar chain structure organized by DNA origami. *J. Am. Chem. Soc.* **132**, 3248–3249 (2010).
11. Mastroianni, A. J., Claridge, S. A. & Alivisatos, A. P. Pyramidal and chiral groupings of gold nanocrystals assembled using DNA scaffolds. *J. Am. Chem. Soc.* **131**, 8455–8459 (2009).
12. Chen, W. *et al.* Nanoparticle superstructures made by polymerase chain reaction: collective interactions of nanoparticles and a new principle for chiral materials. *Nano Lett.* **9**, 2153–2159 (2009).

13. Sharma, J. *et al.* Control of self-assembly of DNA tubules through integration of gold nanoparticles. *Science* **323**, 112–116 (2009).
14. Chen, C.-L. & Rosi, N. L. Preparation of unique 1-D nanoparticle superstructures and tailoring their structural features. *J. Am. Chem. Soc.* **132**, 6902–6903 (2010).
15. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
16. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
17. Fan, Z. & Govorov, A. O. Plasmonic circular dichroism of chiral metal nanoparticle assemblies. *Nano Lett.* **10**, 2580–2587 (2010).
18. Schaaff, T. G. & Whetton, R. L. Giant gold–glutathione cluster compounds: intense optical activity in metal-based transitions. *J. Phys. Chem. B* **104**, 2630–2641 (2000).
19. Shermer, G. *et al.* Chirality of silver nanoparticles synthesized on DNA. *J. Am. Chem. Soc.* **128**, 11006–11007 (2006).
20. George, J. & Thomas, K. G. Surface plasmon coupled circular dichroism of Au nanoparticles on peptide nanotubes. *J. Am. Chem. Soc.* **132**, 2502–2503 (2010).
21. Guerrero-Martínez, A. *et al.* Intense optical activity from three-dimensional chiral ordering of plasmonic nanoantennas. *Angew. Chem. Int. Edn Engl.* **50**, 5499–5503 (2011).
22. Guerrero-Martínez, A., Alonso-Gómez, J. L., Auguié, B., Cid, M. M. & Liz-Marzán, L. M. From individual to collective chirality in metal nanoparticles. *NanoToday* **6**, 381–400 (2011).
23. Berova, N., Nakanishi, K. & Woody, R. W. *Circular Dichroism: Principles and Applications* 2nd edn (Wiley-VCH, 2000).
24. Töring, T., Voigt, N. V., Nangreave, J., Yan, H. & Gotheil, K. V. DNA origami: a quantum leap for self-assembly of complex structures. *Chem. Soc. Rev.* **40**, 5636–5646 (2011).
25. Fan, Z. & Govorov, A. O. Helical metal nanoparticle assemblies with defects: plasmonic chirality and circular dichroism. *J. Phys. Chem. C* **115**, 13254–13261 (2011).
26. Schreiber, R. *et al.* DNA origami-templated growth of arbitrarily shaped metal nanoparticles. *Small* **7**, 1795–1799 (2011).
27. Pilo-Pais, M., Goldberg, S., Samano, E., LaBean, T. H. & Finkelstein, G. Connecting the nanodots: programmable nanofabrication of fused metal shapes on DNA templates. *Nano Lett.* **11**, 3489–3492 (2011).
28. Halas, N. J., Lal, S., Chang, W.-S., Link, S. & Nordlander, P. Plasmons in strongly coupled metallic nanostructures. *Chem. Rev.* **111**, 3913–3961 (2011).
29. Urzhumov, Y. A. *et al.* Plasmonic nanoclusters: a path towards negative-index metafluids. *Opt. Express* **15**, 14129–14145 (2007).
30. Pendry, J. B. A chiral route to negative refraction. *Science* **306**, 1353–1355 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Dietz and G. Acuna for experimental advice and B. Yurke, E. Graugnard, J. O. Rädler and J. P. Kotthaus for discussions. We acknowledge J. Buchner and M. Rief for giving us access to their CD spectrometers, E. Herold for help with the CD measurements, and T. Martin and S. Kempfer for assistance. We also thank D. M. Smith for carefully reading the manuscript. This work was funded by the Volkswagen Foundation, the DFG Cluster of Excellence NIM (Nanosystems Initiative Munich) and the NSF (USA).

Author Contributions A.K., R.S., A.H., F.C.S., A.O.G. and T.L. designed the research. A.K., R.S. and E.-M.R. designed the nanostructures and performed CD measurements. G.P. produced and purified gold samples. A.H. and T.L. investigated ORD effects. Z.F. and A.O.G. performed theoretical calculations. A.K., R.S. and A.O.G. prepared the figures and A.K., R.S., A.H., F.C.S., A.O.G. and T.L. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.L. (tim.liedl@lmu.de).

Asymmetric spiroacetalization catalysed by confined Brønsted acids

Ilija Čorić¹ & Benjamin List¹

Acetals are molecular substructures that contain two oxygen–carbon single bonds at the same carbon atom, and are used in cells to construct carbohydrates and numerous other molecules. A distinctive subgroup are spiroacetals, acetals joining two rings, which occur in a broad range of biologically active compounds, including small insect pheromones and more complex macrocycles^{1,2}. Despite numerous methods for the catalytic asymmetric formation of other commonly occurring stereocentres, there are few approaches that exclusively target the chiral acetal centre and none for spiroacetals^{3,4}. Here we report the design and synthesis of confined Brønsted acids based on a C₂-symmetric imidodiphosphoric acid motif, enabling a catalytic enantioselective spiroacetalization reaction. These rationally constructed Brønsted acids possess an extremely sterically demanding chiral microenvironment, with a single catalytically relevant and geometrically constrained bifunctional active site. Our catalyst design is expected to be of broad utility in catalytic asymmetric reactions involving small and structurally or functionally unbiased substrates.

Spiroacetal natural products are widely found in insects, plants, and bacterial and marine sources, and display a diverse set of biological activities^{1,2}. Their spiroacetal subunit is not only essential for the bioactivity but is also a privileged pharmacophore in drug discovery⁵. Controlling the relative and absolute configuration of spiroacetals can be extremely important, and syntheses of natural products displaying the thermodynamically unfavoured spiroacetal configuration represent daunting challenges². Furthermore, natural products that contain a spiroacetal as the only source of chirality are well known⁶. Indeed, even the parent 6,6-spiroacetal (**1a** in Fig. 1a) is a natural product, olean, the major female-produced sex pheromone of the olive fruitfly (*Bactrocera oleae*). Although isolated as a racemate from natural sources, olean's two mirror images display the remarkable property that one isomer, (*R*)-**1a**, is active on males, and the other one (*S*)-**1a** on females⁷. Asymmetric syntheses of olean have been published but invariably rely on chiral starting materials or reagents⁸. It has been shown that enzymes are capable of stereospecific spiroacetal formation⁹, and organic chemists have developed several powerful synthetic methods for the stereoselective formation of suitably substituted spiroacetals^{2,10–12}. However, so far direct access to spiroacetal motifs in a catalytic enantioselective manner has not been achieved.

Encouraged by our studies of catalytic asymmetric acetalizations³, we targeted chiral Brønsted acid catalysed spiroacetalizations of readily available hydroxyenol ethers, such as **2a** (Fig. 1b, Supplementary Fig. 1). Although the envisaged asymmetric hydroetherification of enol ethers is unknown with chiral Brønsted acid catalysts⁴, achiral acids are routinely used to perform the non-asymmetric reaction. However, an extensive screening of a wide range of known chiral Brønsted acids^{13–18} gave only disappointing results in the generation of both the 5,5- and the 6,6-spiroacetal (the best enantiomeric ratio was 29.5:70.5; Supplementary Tables 1 and 2).

We hypothesized that the absence of a compact chiral environment for the small oxocarbenium ion intermediate, which lacks specific and sterically well defined interactions with its chiral anion, resulted in

poor enantiofacial discrimination. Whereas large active sites can accommodate various transition state geometries leading to different isomers, a confined space could limit this freedom and thereby increase selectivity. On the basis of this hypothesis, we initiated attempts towards modelling sterically demanding acid catalysts for the envisaged spiroacetalization reaction. We realized that a solution to the spiroacetalization problem could potentially also tackle an important issue of broader significance in current Brønsted acid catalysis. Although this field has acquired wide popularity and importance in recent years, reactions of small aliphatic substrates that do not possess sterically demanding protecting groups, large aromatic/planar surfaces, or bulky substituents are still extremely rare^{13–16,19}. We reasoned that this is due to the inability of current synthetic Brønsted acid catalysts to provide a truly compact chiral microenvironment. We speculated that such an environment could effectively restrict molecular motion, leading to efficient enantiodiscrimination even of substrates or intermediates that lack spatially defined interactions, such as hydrogen bonding with the catalyst (Fig. 1c). Here we present the design of simple and readily available synthetic Brønsted acid catalysts featuring a sterically highly demanding and rigid chiral microenvironment around their active site. The design is inspired by enzymes that

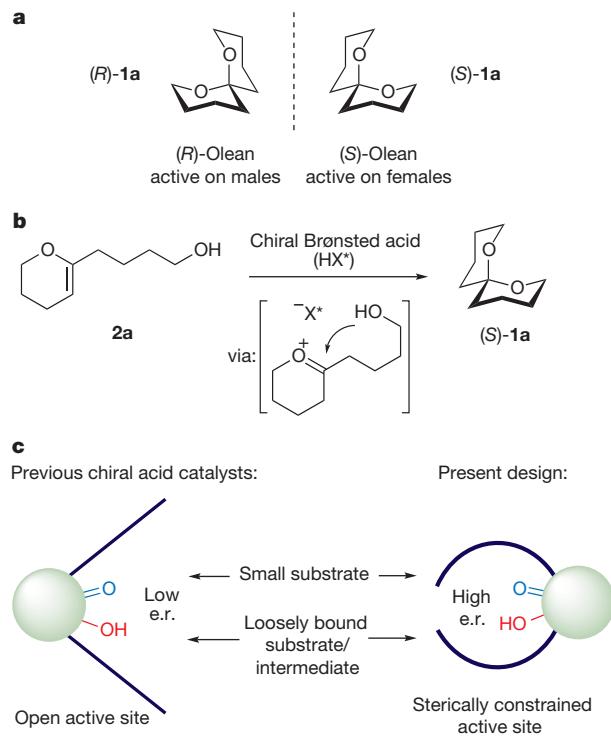


Figure 1 | Towards catalytic asymmetric synthesis of olean. **a**, The two enantiomers of olean from olive fruitfly. **b**, Proposed catalytic asymmetric spiroacetalization. **c**, Catalyst design; see text for details. e.r., enantioselective ratio.

¹Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany.

frequently display deep active site pockets constructed by the folding of the protein, and also by the architectures provided by artificial catalytic cavities^{20,21}.

Our design was based on the analysis of highly successful chiral phosphoric acids, in particular those based on the BINOL (1,1'-binaphthalene-2,2'-diol) backbone^{13–15}, their derivatives, *N*-triflyl phosphoramides¹⁶, and the more recently developed *N*-phosphinyl phosphoramides and *C*₁-symmetric imidodiphosphoric acids²² (Fig. 2a). The active site of a BINOL-derived phosphoric acid is composed of one Brønsted acidic (–OH) and one Brønsted basic site (=O). The remarkable success of phosphoric acids in asymmetric catalysis is widely attributed to bifunctional activation, in which both of these sites are involved in the stabilization of a transition state¹⁵ (Fig. 2b). Although bulky phosphates **3** (Fig. 2a) have found wide application in asymmetric catalysis, it is challenging to further modify their steric environment because 3,3'-substituents on BINOL radiate away from the active site. Recently, significant synthetic efforts have been undertaken by a number of groups to design alternative backbones that would narrow the chiral environment of the phosphoric acid^{23,24}. *C*₁-Symmetric anions **4** and particularly **5** (Fig. 2a) seem to provide a simple alternative way of introducing additional steric demand close to the active site. However, both of these possess several different Brønsted basic sites potentially giving a large number of geometrically distinct acid/base pairs compared to only one such pair present in the phosphoric acid (Fig. 2a, b). This effectively leads to an ensemble of catalytically active species, which can stabilize different transition states, resulting in a low selectivity. Furthermore, rotational freedom around P–N and N–S bonds renders flexible the relative positioning of acidic and basic sites, and therefore the entire chiral environment. Indeed, although anions **4** and **5** display remarkable efficiency in a specific set of reactions, both of these structures lack the generality of phosphoric acids. We reasoned that only one distinct and geometrically fixed acid/base pair, translating into a single type of basic site in the corresponding anion is a prerequisite for the construction of a successful and broadly applicable Brønsted acid catalyst.

On the basis of these considerations, we designed a novel class of Brønsted acids, using a *C*₂-symmetric imidodiphosphate anion (Fig. 2c). In principle, such a *C*₂-symmetric imidodiphosphate moiety has two distinct Brønsted basic sites, O and N. The corresponding acid should have a flexible relative positioning of acid/basic pairs owing to free P–N rotation, and would not satisfy the above mentioned criteria for the design of a general chiral Brønsted acid. However, our catalyst design aimed at restricting the imidodiphosphate moiety to a single *O,O*-*syn* conformation between two identical BINOL subunits with bulky 3,3'-substituents (Fig. 2d). We hypothesized that the inclusion of two BINOL subunits would result in their interlocking due to sterically demanding 3,3'-substituents (Fig. 2d). As a direct consequence, the BINOL subunits are unable to freely rotate and the resulting molecular structure possesses a very high rigidity. Importantly, such an arrangement also resulted in the steric blocking of the undesirable alternative Brønsted basic N-site. As the two BINOL subunits are identical, anion **6** (Fig. 2d) is *C*₂-symmetric, and has therefore only a single type of catalytically relevant Brønsted basic site. Consequently, the corresponding Brønsted acid possesses a single catalytically active bifunctional acid/base pair with a fixed geometry (Fig. 2e). The interlocking of BINOL-subunits could in principle also result in the conformational locking of the imidodiphosphate moiety in the *O,O*-*anti* conformation to give the *O,O*-*anti*-isomer of **6** (Fig. 2c). However, on the basis of our modelling studies we expected that the formation of the corresponding *O,O*-*anti*-isomer will be disfavoured for steric reasons due to the bulky 3,3'-substituents on the backbone.

Following this design, we have synthesized several imidodiphosphoric acids based on anion **6**, including **6a** with 2,4,6-Et₃C₆H₂ substituents on the 3,3'-positions of the two BINOL backbones (Supplementary Table 3). It is noteworthy that imidodiphosphoric acids such as **6a** are easily accessible. Their synthesis requires only a single additional synthetic step compared to the corresponding phosphoric acids (Supplementary Figs 2 and 3). Indeed, the crystal structure of imidodiphosphoric acid **6a** reveals a confined active site deeply buried within a highly sterically demanding chiral environment (Fig. 3a,

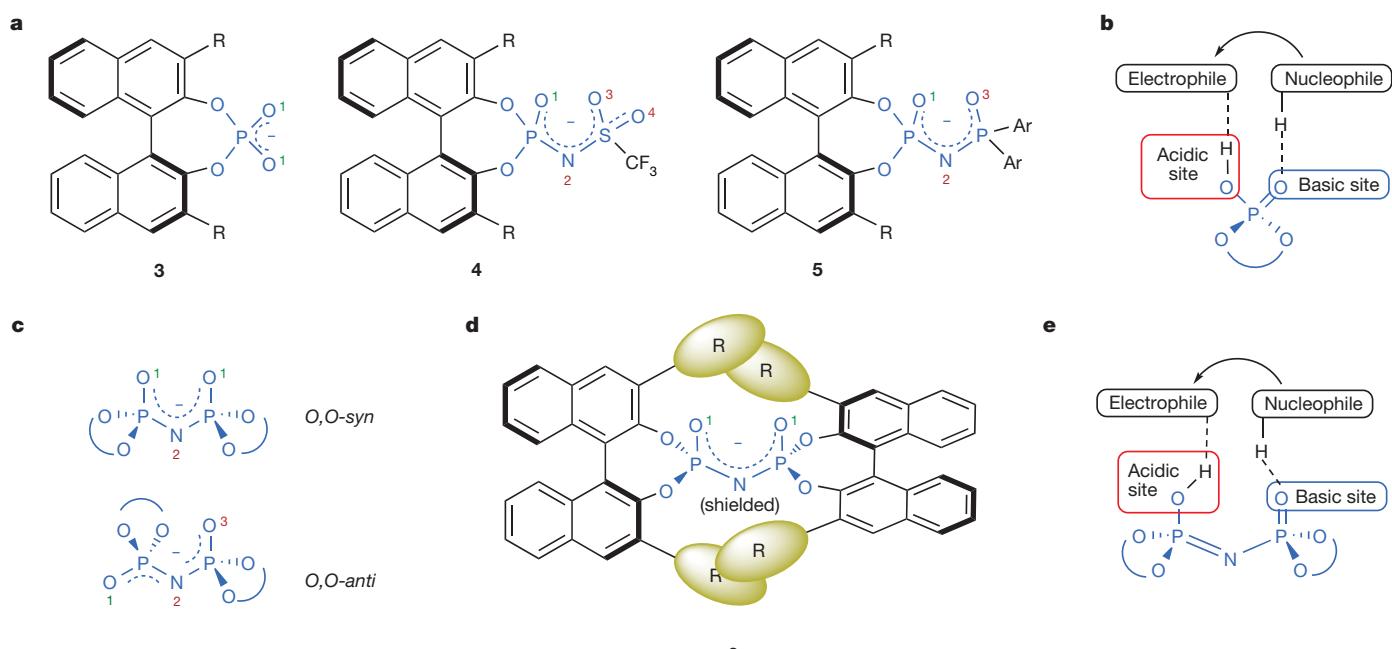


Figure 2 | Development of *C*₂-symmetric imidodiphosphoric acids.

a, Phosphate anion and its derivatives (catalytic moiety shown in blue), showing numbering of different Brønsted basic sites (one in green and additional ones in red). **b**, Bifunctional activation of reactants with Brønsted acid/base pair in phosphoric acid catalysis. **c**, General imidodiphosphate

anion (showing *O,O*-*syn*- and *O,O*-*anti* isomers). **d**, Conformationally locked *O,O*-*syn*-imidodiphosphate anion with *C*₂-symmetry (showing in green bulky 3,3'-substituents responsible for locking). **e**, Bifunctional activation of reactants with Brønsted acid/base pair in imidodiphosphoric acid catalysis.

Supplementary Fig. 4). Viewing the catalytic moiety from the top illustrates how the two BINOL-substituents surround the active site from above (Fig. 3b). Viewed from the bottom, the catalyst demonstrates a tight arrangement of the two remaining BINOL subunits, which completely block access to the imidodiphosphate (Fig. 3c). For comparison, the crystal structure of one of the most popular and sterically demanding phosphoric acids, TRIP (3a, that is, protonated 3 with $R = 2,4,6\text{-}i\text{Pr}_2\text{C}_6\text{H}_2$), is given in Fig. 3d and 3e²⁵. The active site of TRIP is placed inside a broad open cone, whereas that of the imidodiphosphate can better be described as a chiral pocket. Furthermore, the geometrical parameters of the imidodiphosphoric acid, which co-crystallized with a molecule of water, are consistent with a hydronium ion bridging the two imidodiphosphate oxygen atoms (Supplementary Fig. 4, Supplementary Tables 4–6). This supports proton location on oxygen rather than on nitrogen, as shown in Fig. 2e.

Having produced a Brønsted acid confined within an extremely sterically demanding chiral cavity, we turned to the spiroacetalization problem (Supplementary Table 3). Our catalyst design was quickly rewarded with the first (to our knowledge) highly enantioselective catalytic spiroacetalization reaction, obtaining (S)-olean from alcohol 2a with an excellent enantiomeric ratio of 98:2 (Table 1, entry 1). The (R)-enantiomer of olean was easily obtained by using the enantiomer of the catalyst (entry 2). Imidodiphosphoric acid catalyst 6a proved quite general, and various other spiroacetals were obtained with high enantioselectivity. Only 0.1 mol% of catalyst 6a was required in the construction of the small 5,5-spiroacetal (entry 3). The 5,6-spiroacetal motif could be accessed equally successful by cyclization of either five or six-membered rings (entries 4 and 5). Different enol ether ring sizes are similarly tolerated, as demonstrated by highly enantioselective reactions to form the corresponding 7,6- and 7,5-spiroacetals (entries 6 and 7). Notably, all these small spiroacetals are core structures of many natural products^{1,2}.

When other stereocentres are present in the target molecule, accessing the spiroacetal stereocentre in the most stable thermodynamic configuration is usually easy. However, the formation of fragile non-thermodynamic spiroacetals, having the less stable configuration of the spiroacetal stereocentre, presents a formidable challenge for organic synthesis^{2,10,11}. An ideal solution would be provided by a catalyst that could override thermodynamic preference and control the spiroacetal

configuration. Gratifyingly, with our confined Brønsted acid catalyst this goal could be achieved. A variety of non-thermodynamic spiroacetals (1g, 1i, 1k in Table 1) could be formed with diastereomeric ratios ranging from 5:1 to 23:1, against a strong thermodynamic preference of up to 1:124 (entries 8, 10 and 12 in Table 1). Moreover, when the catalyst control matches the inherent thermodynamic preference, spiroacetals can be obtained with excellent diastereoselectivities that can even exceed thermodynamic ratios (entries 9, 11, 13). In this way, thermodynamic spiroacetals can be obtained under very mild, kinetic conditions, as compared to the relatively strongly acidic conditions typically required for thermodynamic equilibration. It is important to note that kinetic spiroacetalizations of substrates 2g–2m (Table 1), catalysed with a simple achiral acid, favoured the thermodynamic product with only poor diastereoselectivity ranging from 1:1 to 1:3. Spiroacetalization of substituted enol ether 2m revealed another powerful capability of the confined Brønsted acid 6a. An efficient kinetic resolution occurred, delivering both bisacetal 1n and enolacetal (R)-2m with excellent enantioselectivity (entry 14). Asymmetric synthesis of these sensitive motifs by any other method would present a considerable challenge. Finally, to further explore the behaviour of our catalyst, we also examined large substrate 2o, an open form of the steroid sapogenin, diosgenin (entry 15). As expected, the reactivity was rather low using either catalyst enantiomer, although only a single diastereomer (the known non-thermodynamic isomer 20-*epi*-diosgenin) was obtained²⁶. This probably results from the inability of our confined acid to efficiently accommodate such large substrates.

Whereas recent developments in Brønsted acid catalysis have enabled an array of transformations to be performed asymmetrically^{13–16,27}, substrates have typically been sterically or electronically biased. We have shown here that confined Brønsted acids with extreme steric demand and chiral pockets reminiscent of those found in enzymes can overcome this limitation and solve an important problem in organic synthesis. Our spiroacetalization reaction offers a straightforward catalytic enantioselective access to spiroacetal motifs, and is expected to find applications in the synthesis of diverse natural products and biologically active molecules. We believe that the concepts described here open the door to the development of asymmetric reactions that include small aliphatic and/or loosely bound molecules, and will be widely applicable, possibly even beyond Brønsted acid catalysis^{28–30}.

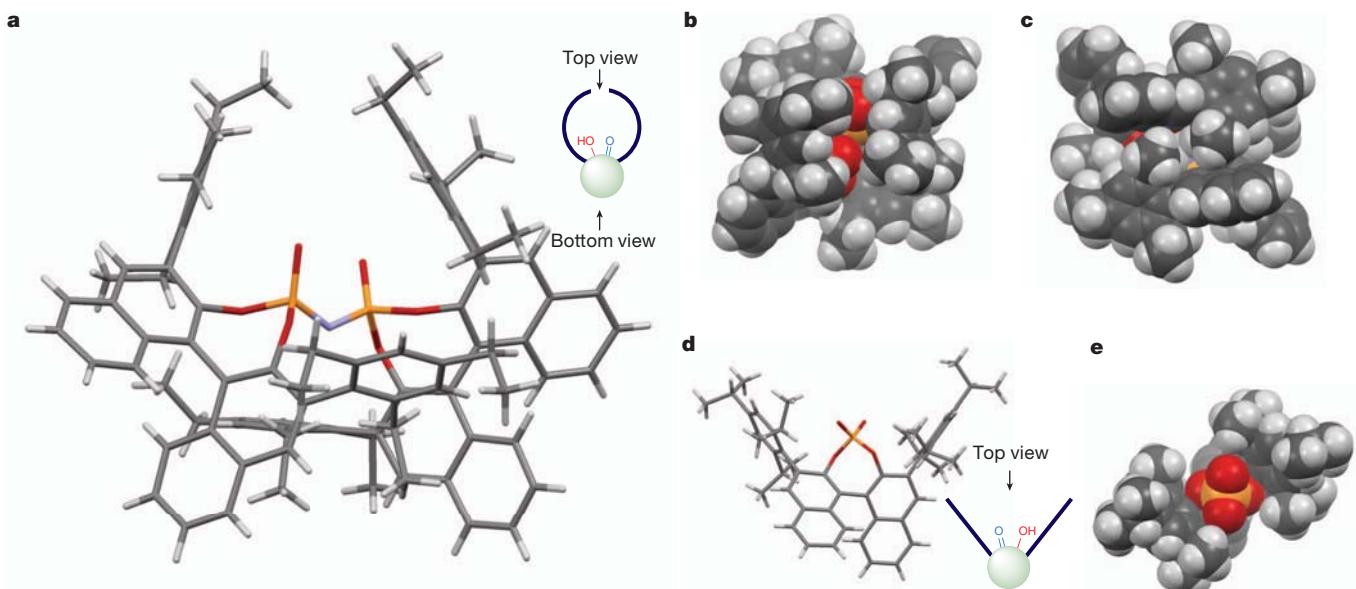
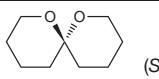
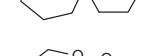
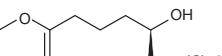
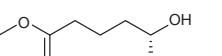
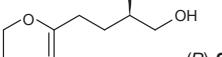
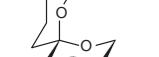
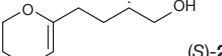
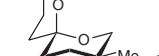
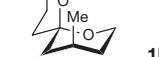
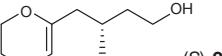
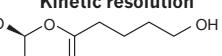
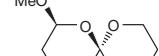
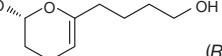
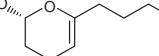
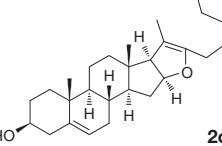
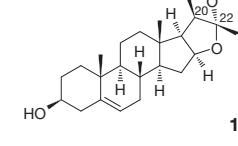


Figure 3 | Comparison of crystal structures of anions of 6a (6 with $R = 2,4,6\text{-}i\text{Pr}_2\text{C}_6\text{H}_2$) and TRIP. a, Crystal structure of the anion of 6a, side view (main figure) and schematic representation (inset). b, 6a, top view at the active site, space filling model. c, 6a, bottom view at the active site, space filling model.

d, Crystal structure of TRIP anion (main figure) and schematic representation (inset). e, TRIP, top view at the active site, space filling model. Colour coding: dark grey, carbon; light grey, hydrogen; red, oxygen; orange, phosphorus; purple, nitrogen.

Table 1 | Catalytic asymmetric spiroacetalization

Entry	Substrate	Reaction details	Product	Yield	e.r. or d.r.
1	 2a	6a (5 mol%), -25°C , MTBE	 (S)-1a	77%	e.r. 98:2
2	 2a	<i>ent</i> - 6a (5 mol%), -25°C , MTBE	 (R)-1a	70%	e.r. 97.5:2.5
3	 2b	6a (0.1 mol%), -55°C , CH_2Cl_2	 1b	62%	e.r. 96:4
4	 2c	6a (1 mol%), -35°C , DCE	 1c	81%	e.r. 95.5:4.5
5	 2d	6a (1 mol%), -35°C , MTBE	 1c	69%	e.r. 96:4
6	 2e	6a (1 mol%), -25°C , MTBE	 1e	78%	e.r. 96:4
7	 2f	6a (1 mol%), -35°C , MTBE	 1f	88%	e.r. 98.5:1.5
8	 (S)-2g		 1g	83%	d.r. 5:1, non-thermodynamic spiroacetal (thermodynamic d.r. 1:124)
9	 (R)-2g	6a (5 mol%), -35°C , MTBE	 1h	89%	d.r. 65:1
10	 (R)-2i	6a (5 mol%), -35°C , MTBE	 1i	70%	d.r. 23:1, non-thermodynamic spiroacetal (thermodynamic d.r. 1:9)
11	 (S)-2i		 1j	70%	d.r. 50:1
12	 (R)-2k		 1k	76%	d.r. 7:1, non-thermodynamic spiroacetal (thermodynamic d.r. 1:60)
13	 (S)-2k	6a (10 mol%), -35°C , MTBE	 1l	86%	d.r. 100:1
14	Kinetic resolution  (S)-2m	6a (5 mol%), -25°C , MTBE (50% conv.)	 1n	70%	e.r. 97.5:2.5 (d.r. >50:1)
	 (R)-2m		 (R)-2m	86%	e.r. 97:3
15	 2o	6a or <i>ent</i> - 6a (1 mol%), 20°C , DCE	 1o	88%	d.r. >20:1:1:1

Reactions were run in the presence of molecular sieves. Substrates in entries 8 to 14 were examined simultaneously as racemic mixtures. The thermodynamic diastereomeric ratio (d.r.) was measured after treating substrates with 3 M HCl(aq) at room temperature. e.r., enantiomeric ratio; MTBE, *tert*-butylmethyl ether; DCE, 1,2-dichloroethane.

METHODS SUMMARY

Catalytic asymmetric synthesis of (S)-olean was performed as follows. MTBE (*tert*-butylmethyl ether) (7 ml) and 4 Å molecular sieves (50 mg) were cooled to -25°C in a vial equipped with a septum. A solution of **2a** (0.25 mmol) in MTBE (2 ml) was added and the mixture stirred for 5–10 min. A solution of catalyst **6a**

(0.0125 mmol, 5 mol%) in MTBE (1 ml) was added dropwise and the mixture was stirred at -25°C for 24 h. Et_3N (50 μl) was added at the same temperature, the mixture concentrated, and the product isolated by column chromatography. For complete experimental details and characterization of compounds, see Supplementary Information.

Received 14 November 2011; accepted 10 February 2012.

1. Perron, F. & Albizati, K. F. Chemistry of spiroketals. *Chem. Rev.* **89**, 1617–1661 (1989).
2. Aho, J. E., Pihko, P. M. & Rissa, T. K. Nonanomeric spiroketals in natural products: structures, sources, and synthetic strategies. *Chem. Rev.* **105**, 4406–4440 (2005).
3. Čorić, I., Vellalath, S. & List, B. Catalytic asymmetric transacetalization. *J. Am. Chem. Soc.* **132**, 8536–8537 (2010).
4. Nagano, H. & Katsuki, T. Stereocontrolled OH protection: asymmetric tetrahydrofuranylation. *Chem. Lett.* **31**, 782–783 (2002).
5. Zinzalla, G., Milroy, L.-G. & Ley, S. V. Chemical variation of natural product-like scaffolds: design and synthesis of spiroketal derivatives. *Org. Biomol. Chem.* **4**, 1977–2002 (2006).
6. Brasholz, M., Sörgel, S., Azap, C. & Reißig, H.-U. Rubromycins: structurally intriguing, biologically valuable, synthetically challenging antitumour antibiotics. *Eur. J. Org. Chem.* 3801–3814 (2007).
7. Haniotakis, G., Francke, W., Mori, K., Redlich, H. & Schurig, V. Sex-specific activity of (*R*)(–)- and (*S*)(+)-1,7-dioxaspiro[5.5]undecane, the major pheromone of *Dacus oleae*. *J. Chem. Ecol.* **12**, 1559–1568 (1986).
8. Redlich, H. & Francke, W. Synthesis of enantiomerically pure 1,7-dioxaspiro[5.5]undecanes, pheromone components of the olive fly (*Dacus oleae*). *Angew. Chem. Int. Ed.* **23**, 519–520 (1984).
9. Takahashi, S. et al. Reveromycin A biosynthesis uses RevG and RevJ for stereospecific spiroacetal formation. *Nature Chem. Biol.* **7**, 461–468 (2011).
10. Takaoka, L. R., Buckmelter, A. J., LaCruz, T. E. & Rychovsky, S. D. Rational synthesis of contra-thermodynamic spiroacetals by reductive cyclizations. *J. Am. Chem. Soc.* **127**, 528–529 (2005).
11. Moilanen, S. B., Potuzak, J. S. & Tan, D. S. Stereocontrolled synthesis of spiroketals via Ti(O-i-Pr)₄-mediated kinetic spirocyclization of glycal epoxides with retention of configuration. *J. Am. Chem. Soc.* **128**, 1792–1793 (2006).
12. Audrain, H., Thorhauge, J., Hazell, R. G. & Jørgensen, K. A. A novel catalytic and highly enantioselective approach for the synthesis of optically active carbohydrate derivatives. *J. Org. Chem.* **65**, 4487–4497 (2000).
13. Akiyama, T., Itoh, J., Yokota, K. & Fuchibe, K. Enantioselective Mannich-type reaction catalyzed by a chiral Brønsted acid. *Angew. Chem. Int. Edn* **43**, 1566–1568 (2004).
14. Uraguchi, D. & Terada, M. Chiral Brønsted acid-catalyzed direct Mannich reactions via electrophilic activation. *J. Am. Chem. Soc.* **126**, 5356–5357 (2004).
15. Akiyama, T. Stronger Brønsted acids. *Chem. Rev.* **107**, 5744–5758 (2007).
16. Nakashima, D. & Yamamoto, H. Design of chiral *N*-triflyl phosphoramide as a strong chiral Brønsted acid and its application to asymmetric Diels–Alder reaction. *J. Am. Chem. Soc.* **128**, 9626–9627 (2006).
17. Doyle, A. G. & Jacobsen, E. N. Small-molecule H-bond donors in asymmetric catalysis. *Chem. Rev.* **107**, 5713–5743 (2007).
18. Huang, Y., Unni, K. A., Thadani, A. N. & Rawal, V. H. Single enantiomers from a chiral-alcohol catalyst. *Nature* **424**, 146 (2003).
19. Zhang, Q.-W. et al. Brønsted acid catalyzed enantioselective semipinacol rearrangement for the synthesis of chiral spiroethers. *Angew. Chem. Int. Edn* **48**, 8572–8574 (2009).
20. Shenoy, S. R., Crisóstomo, F. R. P., Iwasawa, T. & Rebek, J. Jr. Organocatalysis in a synthetic receptor with an inwardly directed carboxylic acid. *J. Am. Chem. Soc.* **130**, 5658–5659 (2008).
21. Hastings, C. J., Pluth, M. D., Bergman, R. G. & Raymond, K. N. Enzymelike catalysis of the Nazarov cyclization by supramolecular encapsulation. *J. Am. Chem. Soc.* **132**, 6938–6940 (2010).
22. Vellalath, S., Čorić, I. & List, B. N-Phosphinyl phosphoramide—a chiral Brønsted acid motif for the direct asymmetric N,O-acetalization of aldehydes. *Angew. Chem. Int. Edn* **49**, 9749–9752 (2010).
23. Xu, F. et al. SPINOL-Derived phosphoric acids: synthesis and application in enantioselective Friedel–Crafts reaction of indoles with imines. *J. Org. Chem.* **75**, 8677–8680 (2010).
24. Čorić, I., Müller, S. & List, B. Kinetic resolution of homoaldols via catalytic asymmetric transacetalization. *J. Am. Chem. Soc.* **132**, 17370–17373 (2010).
25. Klussmann, M. et al. Synthesis of TRIP and analysis of phosphate salt impurities. *Synlett* 2189–2192 (2010).
26. Wall, M. E., Eddy, C. R. & Serota, S. Steroidal saponins. XIX. Stereochemistry of saponins and cholesterol at carbon 20. *J. Am. Chem. Soc.* **76**, 2849–2850 (1954).
27. Shapiro, N. D., Rauniar, V., Hamilton, G. L., Wu, J. & Toste, F. D. Asymmetric additions to dienes catalysed by a dithiophosphoric acid. *Nature* **470**, 245–249 (2011).
28. Mayer, S. & List, B. Asymmetric counteranion-directed catalysis. *Angew. Chem. Int. Edn* **45**, 4193–4195 (2006).
29. Hamilton, G. L., Kang, E. J., Mba, M. & Toste, F. D. A powerful chiral counterion strategy for asymmetric transition metal catalysis. *Science* **317**, 496–499 (2007).
30. Mukherjee, S. & List, B. Chiral counteranions in asymmetric transition-metal catalysis: highly enantioselective Pd/Brønsted acid-catalyzed direct α -allylation of aldehydes. *J. Am. Chem. Soc.* **129**, 11336–11337 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Dreier and R. Goddard for crystal structure analysis of catalyst **6a**. S. Vellalath and S. Müller are acknowledged for donating several previously described catalysts, and N. Wippich and S. Dehn for technical assistance. We gratefully acknowledge support from the Max Planck Society and the European Research Council.

Author Contributions I.Č. and B.L. jointly designed and developed C_2 -symmetric imidodiphosphoric acids, developed the spiroacetalization reaction, and wrote the manuscript. I.Č. conducted the laboratory experiments. B.L. initiated and oversaw the project.

Author Information X-ray crystallographic data have been deposited in the Cambridge Crystallographic Data Centre database (<http://www.ccdc.cam.ac.uk/>) under accession code CCDC 864762. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.L. (list@kofo.mpg.de).

Uncovering the Neoproterozoic carbon cycle

D. T. Johnston¹, F. A. Macdonald¹, B. C. Gill¹, P. F. Hoffman^{1,2} & D. P. Schrag¹

Interpretations of major climatic and biological events in Earth history are, in large part, derived from the stable carbon isotope records of carbonate rocks and sedimentary organic matter^{1,2}. Neoproterozoic carbonate records contain unusual and large negative isotopic anomalies within long periods (10–100 million years) characterized by $\delta^{13}\text{C}$ in carbonate ($\delta^{13}\text{C}_{\text{carb}}$) enriched to more than +5 per mil. Classically, $\delta^{13}\text{C}_{\text{carb}}$ is interpreted as a metric of the relative fraction of carbon buried as organic matter in marine sediments^{2–4}, which can be linked to oxygen accumulation through the stoichiometry of primary production^{3,5}. If a change in the isotopic composition of marine dissolved inorganic carbon is responsible for these excursions, it is expected that records of $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}$ in organic carbon ($\delta^{13}\text{C}_{\text{org}}$) will covary, offset by the fractionation imparted by primary production⁵. The documentation of several Neoproterozoic $\delta^{13}\text{C}_{\text{carb}}$ excursions that are decoupled from $\delta^{13}\text{C}_{\text{org}}$, however, indicates that other mechanisms^{6–8} may account for these excursions. Here we present $\delta^{13}\text{C}$ data from Mongolia, northwest Canada and Namibia that capture multiple large-amplitude (over 10 per mil) negative carbon isotope anomalies, and use these data in a new quantitative mixing model to examine the behaviour of the Neoproterozoic carbon cycle. We find that carbonate and organic carbon isotope data from Mongolia and Canada are tightly coupled through multiple $\delta^{13}\text{C}_{\text{carb}}$ excursions, quantitatively ruling out previously suggested alternative explanations, such as diagenesis^{7,8} or the presence and terminal oxidation of a large marine dissolved organic carbon reservoir⁶. Our data from Namibia, which do not record isotopic covariance, can be explained by simple mixing with a detrital flux of organic matter. We thus interpret $\delta^{13}\text{C}_{\text{carb}}$ anomalies as recording a primary perturbation to the surface carbon cycle. This interpretation requires the revisiting of models linking drastic isotope excursions to deep ocean oxygenation and the opening of environments capable of supporting animals^{9–11}.

There are two leading hypotheses for the documented large-amplitude¹² Neoproterozoic carbon isotope anomalies. The first proposal is that the Neoproterozoic deep ocean carried a massive dissolved organic carbon (DOC) reservoir⁶. This model stems from the observation of unlinked changes in $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ with the postulated large DOC pool allowing for $\delta^{13}\text{C}_{\text{org}}$ records to be buffered against isotopic change. Another view of these large carbon isotopic excursions is represented by a set of hypotheses arguing for the infidelity of $\delta^{13}\text{C}_{\text{carb}}$ records. In this view, secondary alteration by meteoric waters or burial diagenesis is invoked to satisfy carbon isotope decoupling^{7,8}. Given such disparate proposals for explaining the Neoproterozoic carbon isotope record, and the implications for surface environments, climate, oxygen and animals, we aim to revisit the behaviour of the marine carbon cycle and test the validity of the above proposals with a new data set and model.

We sampled stratigraphic sections spanning the mid-Neoproterozoic Cryogenian glacial intermission (<717 to >635 million years ago) at high resolution in Mongolia, northern Namibia and northwest Canada (see Supplementary Information). In Mongolia, limestone of the Tayshir member (of the Tsagaan Oloom Formation) is bracketed by the

Maikhan Ul and Khongoryn diamictites¹³ (Fig. 1). The Tayshir carbon isotope anomaly¹³ is recorded in stratigraphic sections throughout the basin, commences with a distinct flooding surface, and is developed predominantly in limestone micrite with relatively low primary porosity. The Tayshir anomaly can either be correlated to moderately negative values in the Gruis Formation of northern Namibia¹⁴ and Bonahaven Dolomite of the British–Irish Caledonides¹⁵, or to the Trezona anomaly in Australia¹⁶ and Namibia^{17,18}. The Trezona anomaly was also sampled through the Omibaatje Formation in northern Namibia (see Supplementary Information), which directly underlies an erosion surface related to the Ghaub glaciation of 635.5 ± 1.2 million years ago¹⁹. We further sampled the Cryogenian record from between the glaciations in the Mackenzie Mountains of northwest Canada (see Supplementary Information), where mixed carbonate and siliciclastic rocks of the Twitya and Keele formations are bounded by diamictites of the Sturtian-age Rapitan Group²⁰ (716.47 ± 0.24 million years old) and Marinoan-age (~ 635 million years old) Ice Brook Formation. These strata preserve the post-Sturtian Rasthof anomaly in the Twitya Formation and an anomaly at the top of the Keele Formation that can be correlated to the Trezona anomaly²¹. Although there are alternative published correlations for particular anomalies described above, these correlations are not central to our argument.

We present two $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ records through the Tayshir member of Mongolia (Fig. 1). The overall total organic carbon (TOC) contents broadly follow siliciclastic content and are highest at the base of the more distal section (Uliastay Gol), but remain at about 0.08% throughout the succession with no apparent facies dependence (see Supplementary Information). In the Uliastay Gol section ($n = 171$ pairs), an over 10‰ negative $\delta^{13}\text{C}_{\text{carb}}$ anomaly (the Tayshir) at a depth of around 150 m preserves an extremely tight coupling with $\delta^{13}\text{C}_{\text{org}}$ through the entire excursion. Isotopic covariance is observed at even finer stratigraphic scales, such as the inflection recorded at around 50 m depth. Consistent with this, a parallel section 35 km to the east also preserves a tight isotopic coupling and similar net isotopic offset, or $\varepsilon_{\text{TOC}} = \delta^{13}\text{C}_{\text{carb}} - \delta^{13}\text{C}_{\text{org}}$ (Fig. 1; see Supplementary Information).

In northwest Canada, the Cryogenian Twitya and Keele formations provide a comparison to glacial intermission records from Mongolia, and preserve the Rasthof and a latest-Cryogenian isotope excursion (Fig. 2). The post-Sturtian Twitya Formation preserves strong isotopic covariance between carbonate and organic carbon, and an ε_{TOC} value that closely approximates what is extracted from Mongolia (see Supplementary Information). The pre-Marinoan Keele Formation also appears to preserve carbon isotope covariance, but records a more variable ε_{TOC} , ranging from values similar to those observed in Mongolia to that recorded in the Pleistocene epoch (see Supplementary Information)⁵. In contrast to these records of tight covariance is the Trezona anomaly in Namibia (a TOC-poor dolomudstone, dolograinstone and dolomicrobialaminite; see Supplementary Information), in which $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ have little relationship to one another. This later observation is consistent with much of the published Neoproterozoic data^{6,10,11,22}. Our data set thus reflects both the convention (isotopic decoupling) as well as strong evidence of isotopic covariance.

¹Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA. ²School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada.

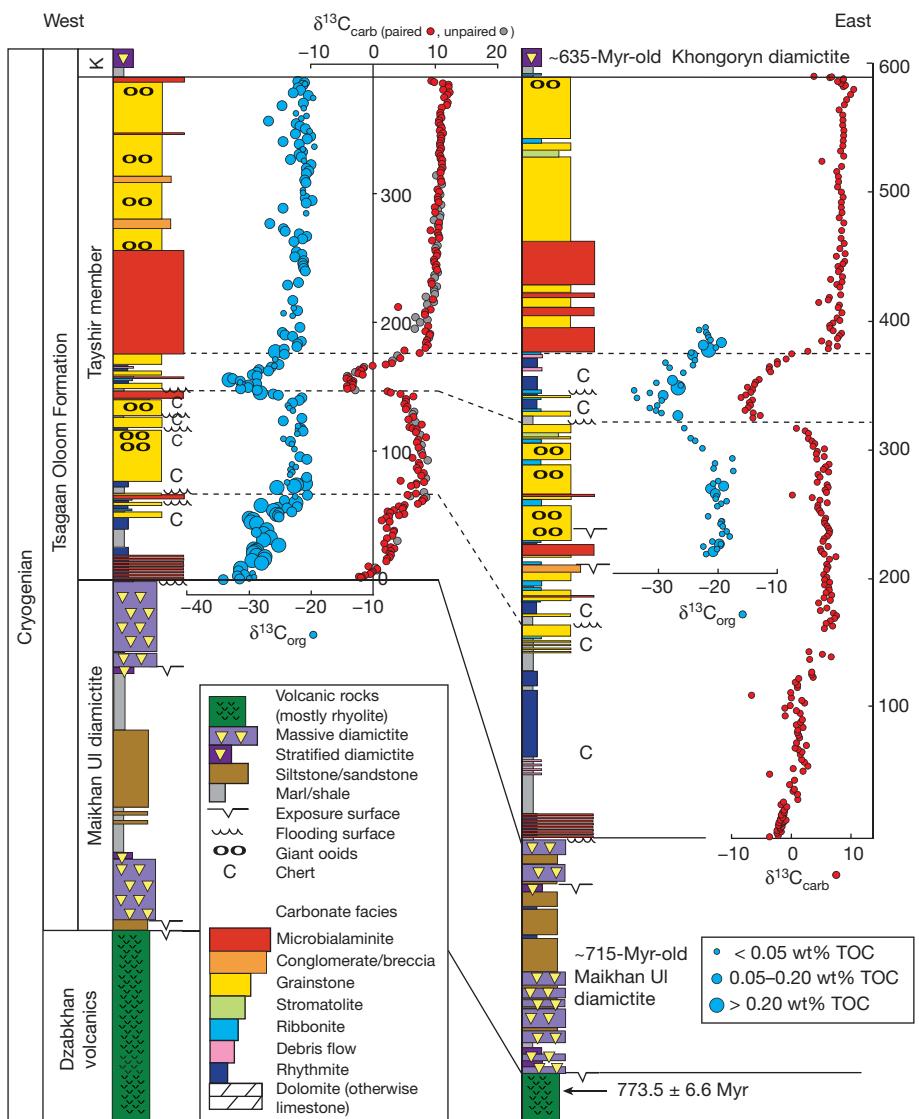


Figure 1 | Chemostratigraphic sections through the Tayshir member in the Uliastay and Tsagaan Gols (river valleys), Mongolia¹³. Red and grey symbols represent $\delta^{13}\text{C}_{\text{carb}}$ data from paired and unpaired samples, respectively (paired $n = 235$). The 2-s.d. errors are smaller than the symbols. The stratigraphic section to the right was measured 35 km to the northeast, a distance shortened by the Cambrian Altai orogeny¹³. Age constraints on the Dzabkhan volcanics are from ref. 31. K, Khongoryn.

We used these data to evaluate the late-Cryogenian carbon cycle. The classic interpretation of isotopic covariance suggests that biomass is synthesized from a relatively well-mixed marine dissolved inorganic carbon (DIC) reservoir and results in an ε_{TOC} value that reflects a kinetic fractionation associated with carbon fixation². This simple depiction of the carbon cycle cleanly explains records from Mongolia and northwest Canada (Figs 1 and 2). Our finding contrasts with models requiring a large DOC pool, which preclude covariance between $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$. That is, data from Mongolia and northwest Canada are quantitatively inconsistent with a large DOC model (Figs 1 and 2). It has also been argued that burial diagenesis⁸ or meteoric alteration⁷ could reset the primary $\delta^{13}\text{C}_{\text{carb}}$, calling into question the robustness of carbonate records and their use as a pre-fossil correlation tool. Rather than targeting the fidelity of $\delta^{13}\text{C}_{\text{org}}$, these studies emphasize the tight covariance between $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{18}\text{O}_{\text{carb}}$ through certain negative $\delta^{13}\text{C}_{\text{carb}}$ excursions as evidence for alteration (see Supplementary Information). However, the observation of a tight coupling between $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ through several Neoproterozoic anomalies (Figs 1 and 2) rules out alteration as the primary cause of all large Neoproterozoic negative $\delta^{13}\text{C}_{\text{carb}}$ excursions (see Supplementary Information). The $\delta^{18}\text{O}_{\text{carb}}$ in these rocks has almost certainly been reset, given fluid-rock exchange and the high concentration of oxygen in water, but explaining how exchange can reset $\delta^{13}\text{C}_{\text{carb}}$ given the enormous mass of carbon in carbonate sequences is more challenging⁸. The correlation of $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{18}\text{O}_{\text{carb}}$ is intriguing and may in fact have an environmental/

diagenetic explanation (see Supplementary Information), but requires further investigation. Importantly, however, diagenesis is not responsible for the excursions through which $\delta^{13}\text{C}$ covaries (Figs 1 and 2).

A complete understanding of the Neoproterozoic carbon cycle requires not only the interpretation of carbon isotope covariance (Figs 1 and 2)^{2,4}, but also an explanation for the data that underpinned these alternative hypotheses. We propose that much of the existing $\delta^{13}\text{C}_{\text{org}}$ from Neoproterozoic sedimentary rocks reflects contamination with a secondary source of organic carbon, possibly from the erosion of organic-rich shale on land or the migration of hydrocarbons within the basin. The isotopic contribution from exogenous sources would be most evident when TOC values are low, consistent with a majority of the published data^{10,22}. This hypothesis does not require that all samples with low TOC carry anomalous $\delta^{13}\text{C}_{\text{org}}$, but notes the greater susceptibility of these samples^{2,4,23} and challenges their universal inclusion in environmental interpretations, especially when $\delta^{13}\text{C}_{\text{org}}$ remains constant through intervals when $\delta^{13}\text{C}_{\text{carb}}$ is variable.

To explore the hypothesis that these data are produced from both an exogenous component and a primary source, we present a model for two-component mixing (Fig. 3a, b). Here, each component carries an organic carbon flux (monitored using TOC) and a distinct $\delta^{13}\text{C}_{\text{org}}$ so the product of mixing does not require a linear relationship between components (see Supplementary Information). One component in the model is organic carbon from primary producers (A in Fig. 3a), with a high mass fraction and an isotopic composition offset from dissolved

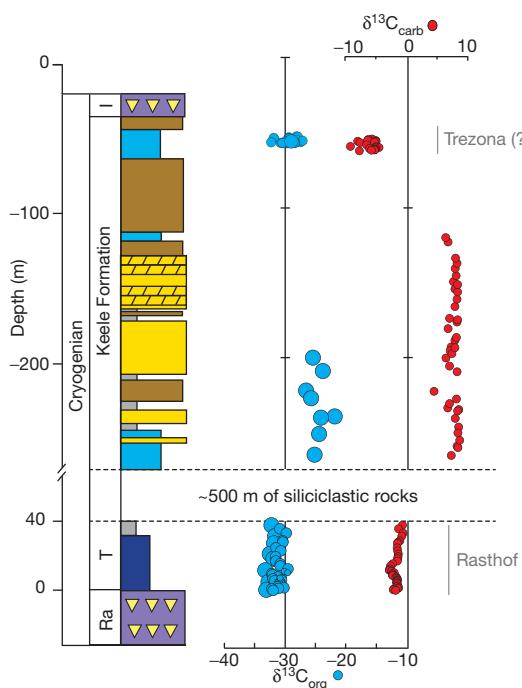
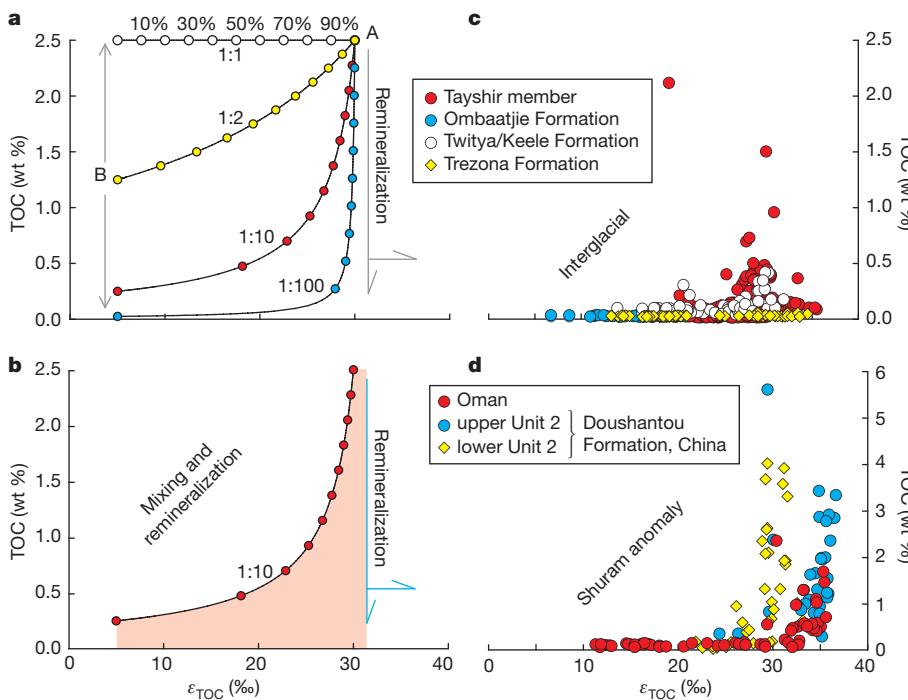


Figure 2 | A chemostratigraphic section through the late Cryogenian period of northwest Canada. As in Fig. 1, red circles represent carbonate carbon and blue denotes paired organic carbon ($n = 73$). Isotope scales are equal for carbonate and organic carbon, and the 2-s.d. errors are smaller than the data points. Stratigraphic keys are as in Fig. 1, as are the TOC (wt%) calibrations. Abbreviations correspond to the following: Ra, Rapitan Group; T, Twitya; I, Ice Brook. Carbonate data in the Keele and Twitya formations is from ref. 21 and overall trends are consistent with ref. 4.

inorganic carbon by ε_{TOC} . The second component (B in Fig. 3a) is exogenous, and could be terrigenous (detrital) inputs, secondary hydrocarbons, or contamination at some point on the rock's life history (with an isotopic composition in the range of typical organic matter). Post-depositional processes, such as heterotrophic remineralization, will also alter mixed contributions (Fig. 3b). When then compared to $\delta^{13}\text{C}_{\text{carb}}$ (Fig. 3c, d), this simple mixing model fully describes



Cryogenian and Ediacaran records without invoking alteration^{7,8} or unfamiliar ocean chemistry^{6,10,11,22}. Our model does not require developing and oxidizing an ocean where DOC overwhelms DIC multiple times throughout the Neoproterozoic era, or driving extensive global alteration of carbonate platform sediments. This does not preclude contributions from DOC²⁴, but the observation of isotopic covariance at the scales observed requires DOC to be significantly subordinate to DIC²⁴. Therefore, we find our model more plausible from a physical and chemical oceanographic standpoint, because it requires only established, well-understood processes (that is, weathering, a circulating ocean and remineralization). It follows from these conclusions that $\delta^{13}\text{C}_{\text{carb}}$ may be a more faithful indicator of environmental change, and $\delta^{13}\text{C}_{\text{org}}$ should be viewed critically, especially when TOC is low and invariant.

Given that $\delta^{13}\text{C}_{\text{carb}}$ may well faithfully preserve primary environmental information, we revisit the reading of this record. The conventional interpretation of the carbon cycle and isotope records uses a simplification of the full carbon flux equation for isotopic mass balance, where $\delta^{13}\text{C}_{\text{vol}} = [\delta^{13}\text{C}_{\text{org}} \times f_{\text{org}}] + [\delta^{13}\text{C}_{\text{carb}} \times (1 - f_{\text{org}})]$, and f_{org} is the relative fraction of carbon buried as organic matter. In the full carbon flux equation, $\delta^{13}\text{C}$ inputs reflect a weighted average of volcanism (carbon entrained during subduction and mantle outgassing) and carbonate and organic carbon weathering^{25,26}, and the sinks include both carbonate and organic matter burial. So if we consider the full carbon flux equation, other ways of driving $\delta^{13}\text{C}$ variations become apparent. Each of the above fluxes can change in magnitude and isotopic composition, and each carry specific consequences for climate and oxidant budgets. Also, the Neoproterozoic $\delta^{13}\text{C}$ excursions occur in tens to hundreds of metres of shallow-water strata (third-order sequence tracts), implying a characteristic timescale of 10^5 to 10^6 years (ref. 27). It is challenging to produce such large (about 10‰) isotopic excursions, which in certain cases plunge beyond the canonical mantle value, on these timescales without the input of an enormous mass of low- $\delta^{13}\text{C}$ carbon, presumably in the form of either organic matter or methane²⁸. This requisite light carbon load may, in part, reflect the mass-balance counterpart to enriched background Cryogenian $\delta^{13}\text{C}_{\text{carb}}$ values. Such an injection of carbon would also significantly raise atmospheric p_{CO_2} or p_{CH_4} and consume existing oxidant reservoirs²⁹, which is difficult to reconcile with the close stratigraphic association of at least some of the excursions and glacial deposits²⁶. Another possibility is that

Figure 3 | A two-component mixing model built to address variability of $\delta^{13}\text{C}$ with changes in TOC. Analytical precision for TOC measurements is 0.05 wt % (1 s.d.). **a**, Where the mass fractions of each component are equal, traditional linear mixing is predicted. As one component increases in relative concentration, the mixing array becomes hyperbolic. The ratios listed on the figure denote the relative contribution of A and B. **b**, An example of the zone of possible solutions (red-shaded region) given contributions from a 1:10 mixing, as well as remineralization and post-deposition processes. These processes can alter samples to lower TOC, with modest changes in $\delta^{13}\text{C}$. In **c** and **d** the data are binned by age. **c**, Interglacial records from Mongolia and northwest Canada, with additional data from the Oombaattie Formation (this study; see Supplementary Information; $n = 79$) and Trezona Formation²². **d**, Published Shuram data from Oman and China^{10,11}. Although the vertical axis is different in **d**, this model approach cleanly explains the excursion.

the oxidation of reduced carbon is accomplished with SO_4^{2-} or Fe oxides as the electron acceptor³⁰, rather than O_2 , which may allow for some stabilization of p_{CO_2} through the effect on alkalinity²⁴. If the duration of these events is found to be longer, then carbon mass input arguments should be modified accordingly, but the relevant reservoirs interacting with the surface carbon cycle remain unchanged.

The data and analysis presented here illustrates that $\delta^{13}\text{C}_{\text{carb}}$ preserves a faithful snapshot of the Neoproterozoic surface carbon cycle. This suggests that there is an environmental explanation for the enriched $\delta^{13}\text{C}_{\text{carb}}$ background values, and that at least some of the 10^5 -to- 10^6 -year negative $\delta^{13}\text{C}_{\text{carb}}$ excursions reflect primary features of the surface carbon cycle. These anomalies thus require the episodic input of isotopically light carbon, presumably methane or organic matter, as well as the cyclic consumption and production of electron acceptors such as sulphate and O_2 . Our data do not eliminate diagenetic explanations for particular excursions throughout the world, but the covariation evident during many of the excursions does suggest a common explanation for all of them. Whether these excursions are driven by the recycling of stored carbon in the form of methane hydrate or ancient organic matter in sediments, or by the dynamics of nutrient-dependent productivity in the ocean remains uncertain. Consistent with both is the idea is that these excursions simply reflect a series of failed transitions from the more reducing and biologically simple Proterozoic world towards the more oxidizing and biologically complex Phanerozoic Earth on which we now live.

METHODS SUMMARY

All isotope data are presented relative to the Vienna PeeDee Belemnite (VPDB) standard and calculated according to standard delta notation and reported in units of ‰. Carbonate carbon and oxygen isotope measurements were made as one individual measurement, with the organic carbon isotopic composition measured separately, but on the same hand sample. Carbonate carbon isotopes were prepared by cutting hand samples to expose a fresh surface followed by micro-drilling 0.001 g. A small aliquot of this powder was analysed on a VG Optima run in dual inlet mode, abutted to a common acid bath preparation device. Samples were acidified in H_3PO_4 at 90 °C. The evolved CO_2 was cryo-focused and analysed against an in-house standard reference gas. The reproducibility of this measurement was better than 0.1‰ (one standard deviation). Organic carbon analyses were performed on large (about 10 g) samples to accommodate low TOC values. Samples were decalcified with concentrated HCl for 48 h, buffered back to a neutral pH (>pH 5), filtered and dried. The mass of insoluble residue was taken as siliciclastic content. Homogenized residues were analysed on a Carlo Erba Elemental Analyzer attached to a ThermoFinnigan Delta V configured in continuous flow mode. Samples and standards were bracketed such that our >350 unique organic carbon analyses (each run in duplicate) were associated with 370 internal standards. These standards have known organic carbon contents and isotope values, and were used to calibrate TOC contents and isotopic compositions. Estimates of error are also reported in the text. The simple mixing model underpinning Fig. 3 is a standard multi-component mixing scenario.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 June 2011; accepted 9 January 2012.

Published online 29 February 2012.

1. Halverson, G. P., Wade, B. P., Hurtgen, M. T. & Barovich, K. M. Neoproterozoic chemostratigraphy. *Precambr. Res.* **182**, 337–350 (2010).
2. Knoll, A. H., Hayes, J. M., Kaufman, A. J., Swett, K. & Lambert, I. B. Secular variation in carbon isotope ratios from upper Proterozoic successions of Svalbard and east Greenland. *Nature* **321**, 832–838 (1986).
3. Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment. *Nature* **359**, 605–609 (1992).
4. Kaufman, A. J., Knoll, A. H. & Narbonne, G. M. Isotopes, ice ages, and terminal Proterozoic earth history. *Proc. Natl Acad. Sci. USA* **94**, 6600–6605 (1997).
5. Hayes, J. M., Strauss, H. & Kaufman, A. J. The abundance of C-13 in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125 (1999).
6. Rothman, D. H., Hayes, J. M. & Summons, R. E. Dynamics of the Neoproterozoic carbon cycle. *Proc. Natl Acad. Sci. USA* **100**, 8124–8129 (2003).

7. Knauth, L. P. & Kennedy, M. J. The late Precambrian greening of the Earth. *Nature* **460**, 728–732 (2009).
8. Derry, L. A. A burial diagenesis origin for the Ediacaran Shuram-Wonoka carbon isotope anomaly. *Earth Planet. Sci. Lett.* **294**, 152–162 (2010).
9. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* **315**, 92–95 (2007).
10. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444**, 744–747 (2006).
11. McFadden, K. A. *et al.* Pulsed oxidation and biological evolution in the Ediacaran Doushantuo Formation. *Proc. Natl Acad. Sci. USA* **105**, 3197–3202 (2008).
12. Grotzinger, J. P., Fike, D. A. & Fischer, W. W. Enigmatic origin of the largest-known carbon isotope excursion in Earth's history. *Nature Geosci.* **4**, 285–292 (2011).
13. Macdonald, F. A., Jones, D. S. & Schrag, D. P. Stratigraphic and tectonic implications of a new glacial diamictite-cap carbonate couplet in southwestern Mongolia. *Geology* **37**, 123–126 (2009).
14. Halverson, G. P., Hoffman, P. F., Schrag, D. P., Maloof, A. C. & Rice, A. H. N. Toward a Neoproterozoic composite carbon-isotope record. *Geol. Soc. Am. Bull.* **117**, 1181–1207 (2005).
15. McCay, G. A., Prave, A. R., Alsop, G. I. & Fallick, A. E. Glacial trinity: Neoproterozoic Earth history within the British-Irish Caledonides. *Geology* **34**, 909–912 (2006).
16. McKirdy, D. M. *et al.* A chemostratigraphic overview of the late Cryogenian interglacial sequence in the Adelaide fold-thrust belt, South Australia. *Precambr. Res.* **106**, 149–186 (2001).
17. Halverson, G. P., Hoffman, P. F., Schrag, D. P. & Kaufman, A. J. A major perturbation of the carbon cycle before the Ghaub glaciation (Neoproterozoic) in Namibia: prelude to snowball Earth? *Geochim. Geophys. Geosyst.* **3**, 1035, <http://dx.doi.org/10.1029/2001GC000244> (2002).
18. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball earth. *Science* **281**, 1342–1346 (1998).
19. Hoffmann, K. H., Condon, D. J., Bowring, S. A. & Crowley, J. L. U-Pb zircon date from the Neoproterozoic Ghaub Formation, Namibia: constraints on Marinoan glaciation. *Geology* **32**, 817–820 (2004).
20. Macdonald, F. A. *et al.* Calibrating the Cryogenian. *Science* **327**, 1241–1243 (2010).
21. Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14**, 129–155 (2002).
22. Swanson-Hysell, N. L. *et al.* Cryogenian glaciation and the onset of carbon-isotope decoupling. *Science* **328**, 608–611 (2010).
23. Dehler, C. M. *et al.* High-resolution delta C-13 stratigraphy of the Chuar Group (ca. 770–742 Ma), Grand Canyon: implications for mid-Neoproterozoic climate change. *Geol. Soc. Am. Bull.* **117**, 32–45 (2005).
24. Tziperman, E., Halevy, I., Johnston, D. T., Knoll, A. H. & Schrag, D. P. Biologically induced Snowball Earth. *Proc. Natl Acad. Sci.* **108** (37), 15091–15096 (2011).
25. Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time. *Phil. Trans. R. Soc. B* **361**, 931–950 (2006).
26. Schrag, D. P., Berner, R. A., Hoffman, P. F. & Halverson, G. P. On the initiation of a snowball Earth. *Geochim. Geophys. Geosyst.* **3**, 1036, <http://dx.doi.org/10.1029/2001gc000219> (2002).
27. Halverson, G. P., Hoffman, P. F., Schrag, D. P. & Kaufman, A. J. A major perturbation of the carbon cycle before the Ghaub glaciation (Neoproterozoic) in Namibia: prelude to snowball Earth? *Geochim. Geophys. Geosyst.* **3**, 1035, <http://dx.doi.org/10.1029/2001gc000244> (2002).
28. Bjerrum, C. J. & Canfield, D. E. Towards a quantitative understanding of the late Neoproterozoic carbon cycle. *Proc. Natl Acad. Sci. USA* **108**, 5542–5547 (2011).
29. Bristow, T. F. & Kennedy, M. J. Carbon isotope excursions and the oxidant budget of the Ediacaran atmosphere and ocean. *Geology* **36**, 863–866 (2008).
30. Johnston, D. T. *et al.* An emerging picture of Neoproterozoic ocean chemistry: Insights from the Chuar Group, Grand Canyon, USA. *Earth Planet. Sci. Lett.* **290**, 64–73 (2010).
31. Levashova, N. M. *et al.* The origin of the Baydaric microcontinent, Mongolia: constraints from paleomagnetism and geochronology. *Tectonophysics* **485**, 306–320 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Laboratory assistance was provided by G. Eischeid, E. Northrop, E. Kennedy, T. O'Brien, A. Breus and A. Masterson. We thank G. Halverson, A. Bradley, E. Tziperman and P. Huybers for discussions and comments. We thank the Yukon Geological Survey, the NSF (grant number EAR-IF 0949227 to D.T.J.), KINSC (Haverford College), Henry and Wendy Breck (to D.P.S.), ESEP (Canadian Institute for Advanced Research, to P.F.H.), Harvard University and NASA NAI (D.T.J. and F.A.M.) for funding.

Author Contributions This project was conceived by D.T.J., F.A.M. and D.P.S. Field work was conducted by F.A.M. and P.F.H. Carbonate carbon analyses were performed by F.A.M. Organic carbon analyses and modelling were carried out by D.T.J. and B.C.G. The paper was written by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.T.J. (johnston@eps.harvard.edu).

METHODS

All isotope data are presented relative to the Vienna PeeDee Belemnite (VPDB) standard and calculated according to standard delta notation and reported in units of ‰. Carbonate carbon and oxygen isotope measurements made as one individual measurement, with the organic carbon isotopic composition measured separately, but on the same hand sample. Carbonate carbon isotopes were prepared by cutting hand samples to expose a fresh surface followed by micro-drilling 0.01 g. A small aliquot of this powder was analysed on a VG Optima run in dual inlet mode, abutted to a common acid bath preparation device. Samples were acidified in H_3PO_4 at 90 °C. The evolved CO_2 was cyro-focused and analysed against an in-house standard reference gas. The reproducibility of this measurement was better than 0.1‰ (one standard deviation). Organic carbon analyses were performed on large (about 10 g) samples to accommodate low TOC values. Samples were decalcified with concentrated HCl for 48 h, buffered back to a neutral pH (>pH 5), filtered, and dried. The mass of insoluble residue was taken as siliciclastic content. Homogenized residues were analysed on a Carlo Erba Elemental Analyzer attached to a ThermoFinnigan Delta V configured in continuous flow mode. Samples and standards were bracketed such that our >350 unique organic carbon analyses (each run in duplicate) were associated with 370

internal standards. These standards have known organic carbon contents and isotope values, and were used to calibrate TOC contents and isotopic compositions. Estimates of error are reported in the text.

The model presented in the manuscript is a multi-component mixing scenario, herein defined by a hyperbolic relationship. The equation represents a style of two-component mixing, where:

$$(\delta^{13}C_{TOC})_M = (\delta^{13}C_{TOC})_A \times f_A \times (X_A/X_M) + (\delta^{13}C_{TOC})_B \times (1 - f_A) \times (X_B/X_M)$$

with the subscripts M, A and B representing the mixture, end-member A and end-member B. There are two weighting factors f_A and $(1 - f_A)$, which are fractional terms seen in the equation describing linear mixing:

$$X_M = X_A \times f_A + X_B \times (1 - f_A)$$

where X is mass. In solving this, we can choose the end-member compositions— X_A , $(\delta^{13}C_{TOC})_A$, X_B and $(\delta^{13}C_{TOC})_B$ —and allow f_A to vary from 0 to 1 (this is a relative flux term). This allows us to solve for X_M (from equation (2) and then $(\delta^{13}C_{TOC})_M$ from equation (1)). For application in Fig. 3, we can include the $\delta^{13}C_{carb}$ on the same sample.

On the difficulty of increasing dental complexity

Enni Harjunmaa¹, Aki Kallonen², Maria Voutilainen¹, Keijo Hääläinen², Marja L. Mikkola¹ & Jukka Jernvall¹

One of the fascinating aspects of the history of life is the apparent increase in morphological complexity through time¹, a well known example being mammalian cheek tooth evolution^{2–4}. In contrast, experimental studies of development more readily show a decrease in complexity, again well exemplified by mammalian teeth, in which tooth crown features called cusps are frequently lost in mutant and transgenic mice^{5–7}. Here we report that mouse tooth complexity can be increased substantially by adjusting multiple signalling pathways simultaneously. We cultured teeth *in vitro* and adjusted ectodysplasin (EDA), activin A and sonic hedgehog (SHH) pathways, all of which are individually required for normal tooth development. We quantified tooth complexity using the number of cusps and a topographic measure of surface complexity⁸. The results show that whereas activation of EDA and activin A signalling, and inhibition of SHH signalling, individually cause subtle to moderate increases in complexity, cusp number is doubled when all three pathways are adjusted in unison. Furthermore, the increase in cusp number does not result from an increase in tooth size, but from an altered primary patterning phase of development. The combination of a lack of complex mutants^{5–7}, the paucity of natural variants with complex phenotypes⁹, and our results of greatly increased dental complexity using multiple pathways, suggests that an increase may be inherently different from a decrease in phenotypic complexity.

Dental complexity, typically measured as the number of tooth crown features such as cusps, is closely linked to an animal's diet. Mammalian species with primarily plant diets have high cheek tooth, or molar, complexity whereas species with animal diets have low complexity, and this pattern seems to hold irrespective of phylogeny^{8,10}. During molar development, cusps form at the positions of epithelial signalling centres, the secondary enamel knots¹¹. Each secondary enamel knot is a replay of the same developmental module, hence each enamel knot expresses the same set of regulatory genes¹¹. This evidence indicates that mammalian molar complexity is developmentally relatively simple and adjusting the dynamics of patterning should readily produce both a decrease and an increase in cusp numbers.

An increasing number of genes are known to affect molar development in the mouse^{5–7}. Of the at least 29 individual gene mutations known, 13 cause the arrest of tooth development, 7 cause a decrease in cusp number, 6 cause modifications in cusp patterns, and 3 cause the formation of one additional cusp (Supplementary Table 1). These patterns contrast with multiple examples of evolutionary increase in cusp number^{2–4,8,10}, and with the paucity of evolutionary decrease in cusp number^{12,13}. Furthermore, both syndromic and normal variation in human dentition predominantly cause simplification or complete loss of teeth^{5,6,14}, agreeing with the data from mice but contrasting with the macroevolutionary patterns.

The prevalence of simple mutants is perhaps to be expected because most mutations may compromise individual signalling pathways required for organ development. Furthermore, the lack of complex mutants could be due to the relatively few gain-of-function mutations studied, or few mutations in feedback inhibitors of major signalling pathways^{5,6} (Supplementary Table 1). This missing 'complexity gene' scenario, however, is not the only plausible hypothesis because a

substantial increase in complexity can also be proposed to require simultaneous changes in several signalling pathways (or multiple changes in a single pathway). This requirement of multiple changes can be conceptualized as 'economics of signalling' in which, for example, increasing signalling through one pathway alone may deplete other signals required to produce a greater number of cusps. In genetic terms, this would approximate a polygenic effect.

To explore the contrasting ideas about increasing complexity, we tested whether adjusting individual signalling pathways alone or in combination could be used to increase tooth cusp number. We cultured mouse molars while adjusting three signalling pathways, EDA, activin A and SHH.

EDA, a diffusible tumour necrosis factor (TNF) family protein, is expressed in the outer epithelium surrounding the forming crown and modulates the activity of enamel knots through its receptor which is expressed in the knot cells¹⁵. Lack of EDA signalling approximately halves the number of cusps in a mouse molar, causing rounder and more closely spaced cusps^{16,17}. Superfluous EDA signalling under the keratin-14 (K14) promoter causes opposite changes in cusp shape and spacing, but only a subtle increase in cusp number¹⁷. Activin β A (also known as inhibin β A) in turn is mesenchymally expressed underneath the enamel knots, and mediates epithelial–mesenchymal interactions required for tooth development¹⁸. Lack of activin A arrests lower molar development before the formation of the primary enamel knot¹⁸. Superfluous activin A applied next to a developing second molar accelerates its development¹⁹, but the effects on cusp numbers are unknown. SHH has been implicated by computational modelling to laterally inhibit formation of enamel knots²⁰, and experimental studies in mouse have shown SHH to inhibit formation of cusps and fusion of adjacent teeth^{21,22}. Whereas *Shh* is initially expressed only in the enamel knots, the SHH pathway is principally active in the surrounding tissue²³. Total lack of SHH signalling causes the arrest of tooth development after the formation of the primary enamel knot²³, but the more subtle effects of SHH on cusp number remain to be quantified.

To tabulate the appearance of enamel knots and cusps in culture conditions, we cultured lower first molar tooth germs of mice that express green fluorescent protein (GFP, fused with Cre-recombinase) in the *Shh* locus (hereafter called *ShhGFP* mice) as previously described^{19,24}. Whereas *Shh* is initially expressed in the enamel knots, during differentiation expression spreads throughout the inner enamel epithelium, thereby the epifluorescence of *ShhGFP* can be used to monitor the development of both cusp number and crown shape (Fig. 1a). First, we determined whether EDA, activin A and SHH can be adjusted individually to increase cusp number. We cultured molars starting at embryonic day 13, just before the primary enamel knot starts to appear. Teeth were treated for 4 days, covering the period of cusp patterning of the first molar, and cusp numbers were tabulated after 8 days in culture (Methods and Supplementary Information).

The results show that the effects of EDA are subtle, lacking apparent increase in cusp number (Fig. 1b). Because the culture conditions limit the detection of phenotypic details, we additionally tested the effects of EDA protein on *Eda*-null mutants which typically develop two to three cusps¹⁷. In these cultures the wild-type cusp number was reached at

¹Developmental Biology Program, Institute of Biotechnology, University of Helsinki, P.O. Box 56, FIN-00014 Helsinki, Finland. ²Division of Materials Physics, Department of Physics, University of Helsinki, P.O. Box 64, FIN-00014 Helsinki, Finland.

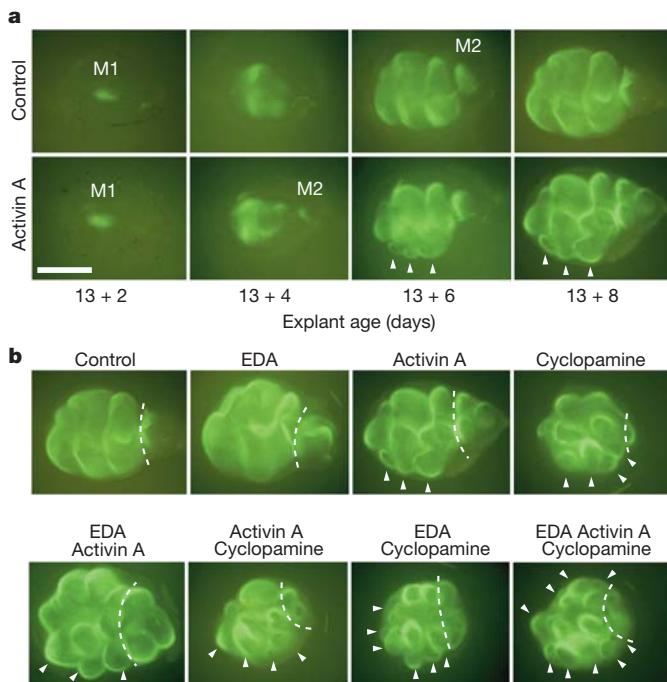


Figure 1 | Cusp number can be increased in cultured teeth. **a**, Culturing *ShhGFP* molar explants from day 13 allows monitoring of the cusp patterning and testing the effects of proteins in the culture medium. For example, activin A causes the appearance of additional cusps lingually (arrowheads). **b**, Culturing teeth with EDA ($0.5 \mu\text{g ml}^{-1}$), activin A ($0.5 \mu\text{g ml}^{-1}$) and cyclopamine ($2 \mu\text{M}$) in the medium causes larger effects on patterning, including additional cusps (arrowheads) when administered in combinations. A dashed line shows the border between the first and the second molars. Teeth were treated for 4 days and cusps were tabulated at 8 days. Anterior is towards the left, buccal towards the top. M1, first molar; M2, second molar. Scale bar, 0.5 mm.

$0.1 \mu\text{g ml}^{-1}$, and higher EDA concentrations increased crown size only (Supplementary Fig. 1 and Supplementary Table 2). Activin A, in contrast, readily caused the formation of additional cusps although high concentrations interfered with growth (Supplementary Fig. 1 and Supplementary Table 2). New cusps were added lingually (Fig. 1), which is also the side that grows more during early molar formation¹¹. The three-row cusp configuration of activin A-treated molars is reminiscent of, for example, molars of multituberculate mammals. For SHH, we first confirmed the polarity of SHH signalling in regulating cusps by culturing teeth with either SHH protein or the alkaloid SHH inhibitor cyclopamine^{25,26}. The results show that SHH stalls and cyclopamine promotes cusp development, hence SHH signalling seems to function as an inhibitor of cusp patterning (Supplementary Fig. 2). Cyclopamine also reduced crown size (Supplementary Fig. 1 and Supplementary Table 2), agreeing with the role of SHH in promoting growth and cell survival^{22,23,27}.

To maximize cusp numbers while minimally impeding growth using EDA, activin A and cyclopamine, we determined the optimum concentrations to be $0.5 \mu\text{g ml}^{-1}$, $0.5 \mu\text{g ml}^{-1}$ and $2 \mu\text{M}$, respectively (Supplementary Table 2 and Fig. 2a). The corresponding average cusp numbers are 5.4, 7.8 and 6.8 (control being 5.6), and changes in crown size are +3%, -17% and -44%, respectively. Next we tested the combined effects of these concentrations of EDA, activin A and cyclopamine on cusp number.

Compared to the single treatments, double-combination cultures showed a greater number of cusps (average cusp numbers are 8.0, 9.0 and 10.1, Figs 1b and 2a), with the triple combination producing the highest number of cusps, effectively doubling the wild-type cusp number (average cusp number is 12.5, Figs 1b and 2a, Supplementary Tables 3 and 4). Additionally, the cyclopamine, double and triple

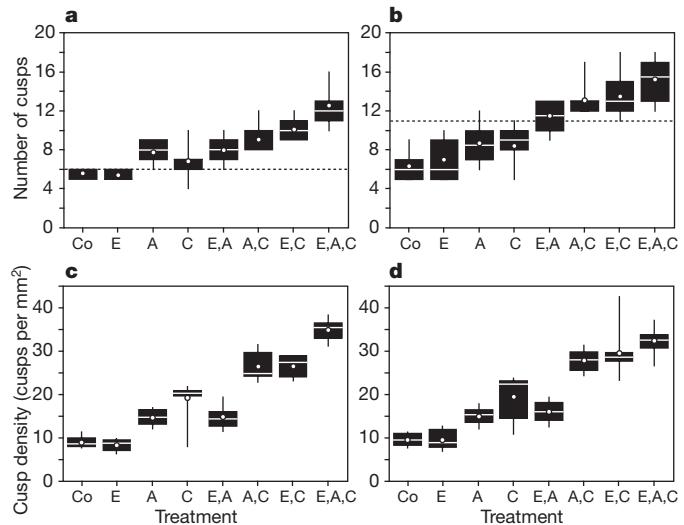


Figure 2 | Number and density of cusps increase with the number of adjusted signalling pathways. **a, b**, Number of cusps calculated for the first molars (a) and combined first and second molars (b). *In vivo* first molar (a) and combined first and second molar (b) cusp numbers are marked with dotted horizontal lines. **c, d**, Cusp densities calculated for the first molars (c) and combined first and second molars (d) show the largest increase using cyclopamine and the smallest increase using EDA. Boxes enclose 50% of observations; the median and mean are indicated with a horizontal bar and circle, respectively, and whiskers denote range. $n = 10$ for each treatment, except (due to uncertain cusp identity) $n = 8$ for A,C and E,C in (a), A,C and E,A,C in (c), and $n = 7$ for E,C in (c). Co, control; E, EDA; A, activin A; C, cyclopamine.

treatments caused fusion of the first and second molars (Fig. 1b). Whereas the daily monitoring of tooth cultures allowed us in most cases (76 out of 80) to identify cusps belonging to each molar, we also tabulated the results using the combined first and second molar cusp numbers (Fig. 2b). Both tabulations show that treatment effects are moderately additive, and correlations between the number of manipulated pathways and the number of cusps are $r_s = 0.79$ for the first molar and $r_s = 0.85$ for the combined first and second molars (both are $P < 0.001$, nonparametric Spearman correlation, Supplementary Fig. 3 and Supplementary Table 5). Even though the second molar does not normally develop well in culture conditions due to inhibition from the first molar¹⁹, the double and triple treatments produced more cusps than in erupted first and second molars combined (Fig. 2b). Furthermore, the triple-treatment molars frequently exceed the cusp number of the complete mouse molar row.

Previously cusp number has been experimentally increased by recombining mouse molar epithelium with the larger molar mesenchyme of the rat²⁸, and by tissue engineering protocols recombining large quantities of mouse dental tissues²⁷. In both cases teeth become large without changing the density of cusps^{27,28}. Because evolutionary increase in cusp number has occurred irrespective of size⁸, and because species differences in cusp number appear during the early patterning phase of development¹¹, we tabulated cusp densities in our cultured teeth. The results show that increases in cusp number are attained through higher packing of cusps, cyclopamine causing the highest increase in cusp density (Fig. 2c, d). Taken together, downregulating SHH signalling decreases inhibition between enamel knots, whereas EDA and especially activin A promote enamel knot formation. These somewhat complementary effects of the three pathways suggest that overcoming bottlenecks in an individual pathway by simultaneously adjusting multiple steps (signalling molecule, receptor and intracellular components) might be less effective in increasing complexity. Nevertheless, our subtle results from EDA-treated molars could be due to downstream bottlenecks, especially because molar growth

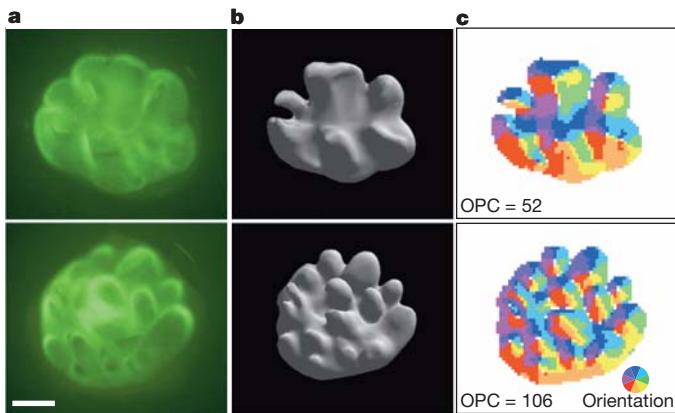


Figure 3 | Tomography reveals high surface complexity of treated molars. **a**, Epifluorescence of *ShhGFP* teeth grown without (top) and with EDA, activin A and cyclopamine (bottom) in culture medium. **b**, Three-dimensional reconstructions of the future enamel–dentin junction from soft tissue X-ray microtomographs reveal a close correspondence to the epifluorescence of the epithelium (**a**). **c**, Slope orientation maps (orientations indicated on the colour wheel) used to calculate OPC values (number of coloured patches). Anterior is towards the left. Scale bar, 0.2 mm.

remains uncompromised. However, when we simultaneously increased EDA and its receptor EDAR by crossing mice overexpressing both genes, the results remain subtle; a distinct new feature being a cingulid, a girdle like structure present in many species but missing in the mouse (Supplementary Fig. 4). These results in turn indicate that EDA affects complexity by fine tuning crown-width-linked features.

The high density of cusps attained raises the question about the maximum possible cusp density. To examine our cultured teeth in detail, we made soft tissue X-ray microtomography scans²⁹ of selected teeth and reconstructed three-dimensional shapes of the future enamel–dentin junctions. The reconstructions show that images of cultured *ShhGFP* teeth are representative of the cusp patterns (Fig. 3a, b). The average cusp diameter of the triple-treatment tooth is 107 μ m (range from 58 to 190 μ m), thereby cusps have roughly 10 cells across in the mesenchyme. In principle a limited number of additional cusps could fit in if the largest cusps became smaller, a change likely to require further changes in signalling.

Finally, we analysed the three-dimensional reconstructions using the orientation patch count (OPC) measure of dental complexity⁸. OPC depicts the number of slopes distinguished by differences in slope orientation irrespective of slope steepness, thereby providing a close correspondence between enamel–dentin junction and tooth surface³⁰. The high cusp count in the triple-treatment tooth is reflected in the high OPC (Fig. 3c), and the crown topography is comparatively similar to the third molar of the giant panda analysed previously⁸. On the basis of OPC values among mammals, including rodents⁸, altering mouse molar patterning produced complexities associated with fibrous plant specializations.

In conclusion, the lack of complex mutants, the high complexity attained adjusting more than one pathway, and the lack of comparable variation in natural populations, all point to a general bias against increase in dental complexity. Yet, the predominant macroevolutionary patterns indicate that ecological factors have been more than adequate in increasing dental complexity. One possible effect of the bias against new cusps could be that, everything else being equal, an increase in dental complexity is slower than a decrease. Of course, yet to be uncovered ‘complexity genes’ may individually increase cusp numbers. Until further evidence, however, we suggest that teeth, and biological structures in general, tend to follow economics of signalling where an increase in complexity beyond the normal variation present in a population requires multiple changes in developmental regulation.

METHODS SUMMARY

Lower molar tooth germs were dissected from heterozygous *ShhGFP* mouse embryos²⁴ at day 13 post fertilization as described previously¹⁹. Explants were photographed daily until day eight of culture, and cusp numbers and identities were followed using the epifluorescence marking the formation of each enamel knot and cusp. We modified the activity of signalling pathways extracellularly, which allowed relatively precise control of timing and dosage in experimental settings. EDA, activin A and SHH signalling pathways were adjusted first individually to determine the concentrations to maximally increase cusp number. Next, teeth were cultured with combinations of EDA, activin A and SHH inhibitor cyclopamine. All treatments were done by adding the ligand or inhibitor into the culturing medium for the first 4 days in culture. Ten molars were used for each treatment and developmental characteristics (cusp number, tooth size and cusp identity) were identified and measured from images using ImageJ 1.43 (<http://rsbweb.nih.gov/ij/>) and analysed using PAST (<http://folk.uio.no/ohammer/past/index.html>). The subtle effects of EDA signalling on crown morphology were tested further by using both *Eda*-null mutants and mice overexpressing *Eda* and its receptor. X-ray tomography, with a custom-built μ CT system Nanotom 180 NF, was used to obtain three-dimensional scans of cusp patterns (with an effective voxel size of 1 μ m). ImageJ was used to manually segment the mesenchyme from the epithelium and ImageJ 3D Viewer plugin was used to render the epithelium–mesenchyme interface and OPC, with 50 data rows in a tooth, was calculated as previously described⁸.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 September 2011; accepted 17 January 2012.

Published online 7 March 2012.

1. Carroll, S. B. Chance and necessity: the evolution of morphological complexity and diversity. *Nature* **409**, 1102–1109 (2001).
2. Hunter, J. P. & Jernvall, J. The hypocone as a key innovation in mammalian evolution. *Proc. Natl Acad. Sci. USA* **92**, 10718–10722 (1995).
3. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs - Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
4. Luo, Z.-X. Transformation and diversification in early mammal evolution. *Nature* **450**, 1011–1019 (2007).
5. Bei, M. Molecular genetics of tooth development. *Curr. Opin. Genet. Dev.* **19**, 504–510 (2009).
6. Nieminen, P. Genetic basis of tooth agenesis. *J. Exp. Zool. B* **312B**, 320–342 (2009).
7. Charles, C. et al. Modulation of *Fgf3* dosage in mouse and men mirrors evolution of mammalian dentition. *Proc. Natl Acad. Sci. USA* **106**, 22364–22368 (2009).
8. Evans, A. R., Wilson, G. P., Fortelius, M. & Jernvall, J. High-level similarity of dentitions in carnivorans and rodents. *Nature* **445**, 78–81 (2007).
9. Miles, A. E. W. & Grigson, C. *Colyer's Variations and Diseases of the Teeth of Animals* (Cambridge Univ. Press, 2003).
10. Santana, S. E., Strait, S. & Dumont, E. R. The better to eat you with: functional correlates of tooth structure in bats. *Funct. Ecol.* **25**, 839–847 (2011).
11. Jernvall, J., Keränen, S. V. E. & Thesleff, I. Evolutionary modification of development in mammalian teeth: quantifying gene expression patterns and topography. *Proc. Natl Acad. Sci. USA* **97**, 14444–14448 (2000).
12. Kay, R. F. “Giant” tamarin from the Miocene of Colombia. *Am. J. Phys. Anthropol.* **95**, 333–353 (1994).
13. Uhen, M. D. Evolution of marine mammals: back to the sea after 300 million years. *Anat. Rec.* **290**, 514–522 (2007).
14. Gábris, K., Fabián, G., Kaán, M., Rózsa, N. & Tarján, I. Prevalence of hypodontia and hyperdontia in paedodontic and orthodontic patients in Budapest. *Community Dent. Health* **23**, 80–82 (2006).
15. Mikkola, M. L. TNF superfamily in skin appendage development. *Cytokine Growth Factor Rev.* **19**, 219–230 (2008).
16. Grünberg, H. Genes and genotypes affecting the teeth of the mouse. *J. Embryol. Exp. Morphol.* **14**, 137–159 (1965).
17. Kangas, A. T., Evans, A. R., Thesleff, I. & Jernvall, J. Nonindependence of mammalian dental characters. *Nature* **432**, 211–214 (2004).
18. Ferguson, C. A. et al. Activin is an essential early mesenchymal signal in tooth development that is required for patterning of the murine dentition. *Genes Dev.* **12**, 2636–2649 (1998).
19. Kavanagh, K. D., Evans, A. R. & Jernvall, J. Predicting evolutionary patterns of mammalian teeth from development. *Nature* **449**, 427–432 (2007).
20. Salazar-Ciudad, I. & Jernvall, J. A computational model of teeth and the developmental origins of morphological variation. *Nature* **464**, 583–586 (2010).
21. Cho, S.-W. et al. Interactions between Shh, Sostdc1 and Wnt signalling and a new feedback loop for spatial patterning of the teeth. *Development* **138**, 1807–1816 (2011).
22. Ahn, Y., Sanderson, B. W., Klein, O. D. & Krumlauf, R. Inhibition of Wnt signalling by Wise (Sostdc1) and negative feedback from Shh controls tooth number and patterning. *Development* **137**, 3221–3231 (2010).

23. Dassule, H. R., Lewis, P., Bei, M., Maas, R. & McMahon, A. P. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development* **127**, 4775–4785 (2000).
24. Harfe, B. D. *et al.* Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
25. Taipale, J. *et al.* Effects of oncogenic mutations in *Smoothened* and *Patched* can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).
26. Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signalling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16**, 2743–2748 (2002).
27. Ishida, K. *et al.* The regulation of tooth morphogenesis is associated with epithelial cell proliferation and the expression of Sonic hedgehog through epithelial–mesenchymal interactions. *Biochem. Biophys. Res. Commun.* **405**, 455–461 (2011).
28. Cai, J. *et al.* Patterning the size and number of tooth and its cusps. *Dev. Biol.* **304**, 499–507 (2007).
29. Metscher, B. D. MicroCT for developmental biology: a versatile tool for high-contrast 3D imaging at histological resolutions. *Dev. Dyn.* **238**, 632–640 (2009).
30. Skinner, M. M. *et al.* Brief communication: contributions of enamel-dentine junction shape and enamel deposition to primate molar crown complexity. *Am. J. Phys. Anthropol.* **142**, 157–163 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Thesleff, P. Munne, A. R. Evans, I. Corfe, J. Moustakas, M. Murtoniemi, I. Salazar-Ciudad, S. Sova, J.-P. Suuronen and S. Zohdy for discussions or help; R. Santalahti, R. Savolainen and M. Mäkinen for technical assistance; M. Hyvönen for the activin A protein; P. Schneider for the Fc-EDA-A1-protein; and C. Tabin and A. Gritli-Linde for the *ShhGFP* mice. This study was funded by the Academy of Finland, the Sigrid Juselius Foundation, the Finnish Cultural Foundation, and the graduate school GSBM.

Author Contributions E.H. and J.J. designed the study. E.H. performed developmental experiments and measurements. M.L.M. and M.V. designed and performed *Eda;Edar* transgenic mouse experiments. K.H. and A.K. designed and performed microtomography imaging. E.H. and J.J. analysed the data and wrote the manuscript with contributions from the other authors. J.J. coordinated the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.J. (jernvall@fastmail.fm).

METHODS

Molar cultures. Lower first molar germs were dissected from heterozygous *ShhGFP* mouse embryos²⁴ at day 13 post fertilization (day 13) and cultured at 37 °C with 5% CO₂ using a Trowell type organ culture as described previously³¹. Molar germs were separated from the jaw tissue, which if left in place, would grow and stunt the development of molars in culture. Media was replaced every 2 days and ascorbic acid was added (100 µg ml⁻¹, Sigma-Aldrich). The epifluorescence of *ShhGFP* was used to select only molars that were at a similar stage of development for treatments. Additionally *ShhGFP* mice were crossed with *Eda*-null mice (Stock 000314, Jackson Laboratory).

Molar treatments. We modified the activity of signalling pathways extracellularly because it allowed quantitative manipulations. EDA, activin A and SHH are examples of different pathways regulating tooth development, and their choice for the experiments was based on previous results on their effects on patterning, theoretical models, and the availability of suitable molecules to adjust signalling activity. All treatments were for the first 4 days, after which the first molar patterning is mostly complete, and the second molar is partially complete. We note that our strategy of maximising cusp number with a single factor (the ligand) in each pathway could in principle be surpassed by also adjusting intracellular components, but these alterations would qualify as multiple changes in developmental regulation.

The phenotypic effects of recombinant Fc-EDA-A1 (ref. 32), activin A³³, SHH (catalogue no. 1314-SH, R&D Systems) proteins and cyclopamine (catalogue no. C4116, Sigma) were tested using different concentrations. Dilutions were made into bovine serum albumin 0.1% (BSA, Sigma) in PBS for the proteins and into dimethylsulphoxide (DMSO) for cyclopamine. Tested EDA concentrations were 0.5 and 0.75 µg ml⁻¹ using wild-type teeth, and 0.01, 0.1, 0.4 and 1.0 µg ml⁻¹ using *Eda*-null mutant teeth. Tested activin A concentrations were 0.1, 0.2, 0.5, 0.75, 3 and 5 µg ml⁻¹. Tested SHH concentrations were 3 and 5 µg ml⁻¹ and cyclopamine concentrations were 0.5, 1, 2 and 4 µM. Only the 3 µg ml⁻¹ concentration seems to be informative for SHH in tooth cultures because the 5 µg ml⁻¹ concentration stopped development, and lower SHH concentrations had no visible effects (E.H. and J.J., unpublished observations). Because high dosages of recombinant proteins may compromise tissue growth, either directly or through impurities, in the Fig. 2 experiments we used concentrations that maximally increased cusps while minimally impeding growth. We note that we cultured heterozygous *ShhGFP* molars because the *ShhGFP* construct is a *Shh* null allele. Although previous studies have reported the development and morphology of heterozygous *ShhGFP* mice to be normal^{19,24}, we tested the effects of SHH and cyclopamine on wild-type molars and the results remained essentially the same (Supplementary Fig. 2).

To examine the subtle effects of EDA signalling on crown morphology, *Eda* and its receptor *Edar* were simultaneously overexpressed in transgenic mice under the K14 promoter. To obtain compound K14-*Eda*; K14-*Edar* transgenic mice,

K14-*Eda* mice¹⁷ were crossed with K14-*Edar* mice³⁴ and teeth of adult mice were analysed using X-ray tomography (using 3 µm voxel size).

Quantitative analyses of experimental data. Cultures were photographed daily using both visible light (Olympus SZX9) and fluorescent (Leica MZFLIII) microscopy. On the last day of culturing, extra tissue surrounding the molar was removed before photography. Ten unpaired molars from at least three different litters were used for each treatment, thus a total of 80 teeth and 1,280 images were analysed for the cusp number tabulations. Developmental characteristics (cusp number, tooth size and cusp identity, Supplementary Table 3) were identified and measured from images using ImageJ 1.43 (<http://rsbweb.nih.gov/ij/>) and analysed using PAST (<http://folk.uio.no/ohammer/past/index.html>). Reliable identification of the two merged cusps forming the anteroconid of the first mouse molar is difficult in cultured teeth, and we therefore tabulated the anteroconid, when present, as one cusp. **X-ray microtomography and OPC analyses.** To examine cusp patterns of the relatively large teeth starting to form hard tissues, we used X-ray tomography to reconstruct three-dimensional cusp patterns. Molars were fixed with 4% paraformaldehyde (PFA), dehydrated into 70% ethanol, and dyed with phosphotungstic acid (catalogue no. P4006, Sigma) for 24 h²⁹. The samples were scanned using a custom-built µCT system Nanotom 180 NF (phoenixX-ray Systems + Services GmbH) with a CMOS detector (Hamamatsu Photonics) and a high-power transmission-type X-ray nanofocus source with a tungsten anode. The samples were imaged with 80 kV acceleration voltage and 180 µA tube current. Projection images were acquired over a full circle of rotation with 0.25° angular intervals, and each projection image was composed of the average of 10 transmission images with 750 ms exposure time. The measurement geometry resulted in an effective voxel size of 1 µm. The reconstruction from the projection images was performed with software datos|x rec supplied by the system manufacturer. Mesenchyme was segmented from epithelium after down-sampling (to 2 µm resolution) using ImageJ (Supplementary Fig. 5). ImageJ 3D Viewer plugin was used to render the epithelium–mesenchyme interface which was cleaned (0.5 cluster decimation and two steps of Laplacian smoothing) in MeshLab 1.3.0 (<http://meshlab.sourceforge.net/>). From the digital elevation models (DEMs) we calculated OPC as previously described⁸. To obtain tooth OPC values comparable to previous studies, the DEMs were down-sampled into a grid having 50 data rows along the tooth length. Reconstructions of adult mouse teeth were visualised using Avizo Fire (Visualization Sciences Group).

31. Närhi, K. & Thesleff, I. Explant culture of embryonic craniofacial tissues: analyzing effects of signaling molecules on gene expression. *Methods Mol. Biol.* **666**, 253–267 (2010).
32. Gaide, O. & Schneider, P. Permanent correction of an inherited ectodermal dysplasia with recombinant EDA. *Nature Med.* **9**, 614–618 (2003).
33. Harrington, A. E. *et al.* Structural basis for the inhibition of activin signalling by follistatin. *EMBO J.* **25**, 1035–1045 (2006).
34. Pispal, J. *et al.* Tooth patterning and enamel formation can be manipulated by misexpression of TNF receptor Edar. *Dev. Dyn.* **231**, 432–440 (2004).

Coevolution in multidimensional trait space favours escape from parasites and pathogens

R. Tucker Gilman¹, Scott L. Nuismer² & Dwueng-Chwuan Jhwueng^{1,†}

Almost all species are subject to continuous attack by parasites and pathogens. Because parasites and pathogens tend to have shorter generation times^{1,2} and often experience stronger selection due to interaction than their victims do^{3,4}, it is frequently argued that they should evolve more rapidly and thus maintain an advantage in the evolutionary race between defence and counter-defence^{4,5}. This prediction generates an apparent paradox: how do victim species survive and even thrive in the face of a continuous onslaught of more rapidly evolving enemies⁵? One potential explanation is that defence is physiologically, mechanically or behaviourally easier than attack, so that evolution is less constrained for victims than for parasites or pathogens⁶. Another possible explanation is that parasites and pathogens have enemies themselves and that victim species persist because parasites and pathogens are regulated from the top down and thus generally have only modest demographic impacts on victim populations^{7,8}. Here we explore a third possibility: that victim species are not as evolutionarily impotent as conventional wisdom holds, but instead have unique evolutionary advantages that help to level the playing field. We use quantitative genetic analysis and individual-based simulations to show that victims can achieve such an advantage when coevolution involves multiple traits in both the host and the parasite.

Most coevolutionary theory has focused on cases in which the interaction rate between two species is governed by a single trait in each species⁹ (but see refs 10–12). In nature, however, species–species interaction rates can depend on multiple traits¹³. For example, the resistance of wild parsnip to webworm attack depends on flowering phenology and on the concentrations of at least two different furanocoumarins in the plant¹⁴. Similarly, many teleost fish use both mucosal barriers and biocidal secretions to defend against monogenean and copepod parasites, and parasites must overcome both mechanisms to infect the host successfully¹⁵. Recent work has shown that the number of traits under selection can have qualitative effects on the outcome of evolution in single-species systems¹⁶. Thus, it is reasonable to ask whether the number of traits may also affect the outcome of coevolution.

To investigate the influence of the number of traits on coevolutionary trajectories, we developed a model of a victim–exploiter system in which the probability of a successful attack by an exploiter on a victim depends on n traits in each species. We assumed that each trait is functionally paired with a trait in the opposite species. When an exploiter encounters a potential victim, the probability of a successful attack depends on the suites of trait values in each individual. We call the probability of successful attack for a given encounter the ‘interaction probability’ to distinguish it from the interaction rate, which also depends on the densities of both species. The interaction probability may be highest when the exploiter’s trait value exceeds that of the victim (that is, difference traits) or when the individuals have similar trait values (that is, matching traits) at each functional pair. For example, the pericarp thickness of the Japanese camellia (*Camellia japonica*) and the rostrum length of its seed-parasitic weevil (*Curculio camelliae*) comprise a difference trait pair¹⁷, whereas the flowering phenology of *Heuchera*

grossularifolia and the emergence time of its seed-parasitic moth *Greyia politella* comprise a matching trait pair¹⁸. Differences in interaction probabilities between individuals with different trait values impose selection on traits and lead to coevolution between the species. We assumed that each trait affects the interaction probability independently. Thus, the victim can achieve a low interaction probability if it evolves at least one defence mechanism that avoids or overcomes the exploiter’s attack. Biologically, this corresponds to the common situation in which an attacker must overcome all of the chemical, morphological, physiological and behavioural defences of the victim for its attack to succeed^{13,15,19}. If the victim can evolve to lower its interaction probability with its exploiter, we say the victim ‘wins’ the evolutionary contest.

We used a quantitative genetic analysis^{20,21} of our model to extend Kirkpatrick’s concept of maximum evolvability²² for a single species to a maximum evolutionary escape rate for a victim–exploiter system (Methods). When the maximum evolutionary escape rate is greater than zero, the victim evolves to overcome or avoid the exploiter’s attack, and so wins the evolutionary contest. Our analysis assumes fixed genetic variance–covariance matrices (that is, G-matrices²⁰) in each species and weak selection due to interspecific interactions. For a very general class of interaction probability functions in the difference trait case, the maximum evolutionary escape rate is

$$\delta_{d,\max} = \max[-(s_v \mathbf{G}_v + s_e \mathbf{G}_e) \boldsymbol{\xi}_a], \quad (1)$$

where \mathbf{G}_v and \mathbf{G}_e are the G-matrices of the victim and exploiter populations, s_v and s_e represent the effect of one interaction on the relative fitness of a victim or an exploiter, and $\boldsymbol{\xi}_a$ is a vector that captures the strength of effect of traits 1 to n on the interaction probability. For a very general class of interaction probability functions in the matching trait case, the maximum evolutionary escape rate is

$$\delta_{m,\max} = \max[\text{eig}[-(s_v \mathbf{G}_v + s_e \mathbf{G}_e) \boldsymbol{\xi}_{\alpha}]], \quad (2)$$

where $\boldsymbol{\xi}_{\alpha}$ is a diagonal matrix in which the i th entry captures the strength of effect of trait i on the interaction probability.

In representative systems where victim and exploiter G-matrices were randomly and independently generated, increasing the number of traits or the strength of correlations between traits increased the probability that the maximum evolutionary escape rate would be greater than zero, and thus that the victim would win the evolutionary contest (Fig. 1). Intuitively, this is true because victims need to overcome their exploiters at only one trait if they are to escape, whereas exploiters must overcome their victims at all traits if they are to succeed. Each additional trait that affects the interaction probability provides the victim with an additional opportunity to evolve an effective escape mechanism. Correlations between traits constrain the evolution of some traits but facilitate the evolution of others. When correlations are strong, it is more likely that the victim’s evolution will be strongly facilitated or the exploiter’s evolution will be strongly constrained in at least one trait, and thus that the victim will evolve to escape the exploiter at that trait.

Our quantitative genetic analysis assumes that selection is weak and G-matrices are constant, but in nature selection due to interspecific

¹National Institute for Mathematical and Biological Synthesis, Knoxville, Tennessee 37916, USA. ²Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, Idaho 83844, USA. [†]Present address: Department of Statistics, Feng-Chia University, Taichung, 40724, Taiwan.

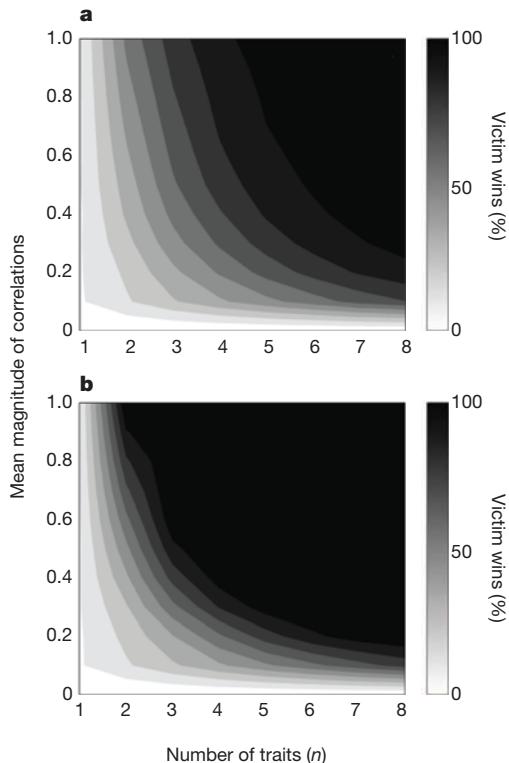


Figure 1 | Percentage victim wins as a function of the number of traits n and the mean magnitude of correlations between traits. **a**, The difference trait case. **b**, The matching trait case. In all simulations $s_e = 1$ and $s_v = -0.3$. In the difference trait case $\xi_{ai} = \xi_{aj}$ and in the matching trait case $\xi_{x,ii} = \xi_{x,jj}$ for all $[i,j]$.

interactions can be strong and G-matrices can evolve^{23–25}. We used genetically explicit individual-based simulations^{24,25} to test the robustness of our analytical results when the assumptions of our quantitative genetic analysis were violated (Methods). We initially assumed that the sizes of the victim and exploiter populations were independent of the interspecific interaction rate. This might be true, for example, if the exploiter has alternative victim species and if the victim population is regulated by something other than exploitation (for example, resource competition). As predicted, victims were more likely to evolve to lower their interaction probabilities with their exploiters when the number of traits governing the interaction probability was large and when the correlations between traits were strong (Tables 1 and 2). This result holds whether interaction probabilities are governed by difference traits, matching traits, or combinations of difference and matching traits (Methods and Supplementary Table 1).

The mechanisms that promote evolutionary escape in the quantitative genetic analysis also operate in numerical simulations. However, numerical simulations revealed an additional mechanism: G-matrices in the victim and exploiter populations evolve to favour the victim

Table 1 | Outcome of coevolution in the difference trait case

n	$r_p = 0.00$	$r_p = 0.08$	$r_p = 0.26$	$r_p = 1.00$
1	0.7222* (17*)	– (–)	– (–)	– (–)
2	0.6703 (17)	0.6295 (18)	0.5963 (32)	0.4621 (39)
4	0.5633 (19)	0.5267 (31)	0.3304 (63)	0.1745 (77)
8	0.4639 (26)	0.4024 (34)	0.1789 (82)	0.0336 (96)

We conducted 100 simulations for each combination of trait number n and correlation strength r_p . We recorded a victim win if the mean interaction probability decreased over 2,000 generations of interaction. Results are shown as mean final interaction probability (percentage victim wins). The parameter r_p is the mean magnitude of correlations between pairs of traits in the same species at the beginning of each simulation. Biologically, r_p can be interpreted as the expected magnitude of correlations between traits due to pleiotropy alone. The magnitude and direction of correlations can change as G-matrices evolve (see Methods), and we measured r_p at the beginning of each simulation to capture correlation strength before selection had altered G-matrices.

*Correlation is not meaningful when $n = 1$.

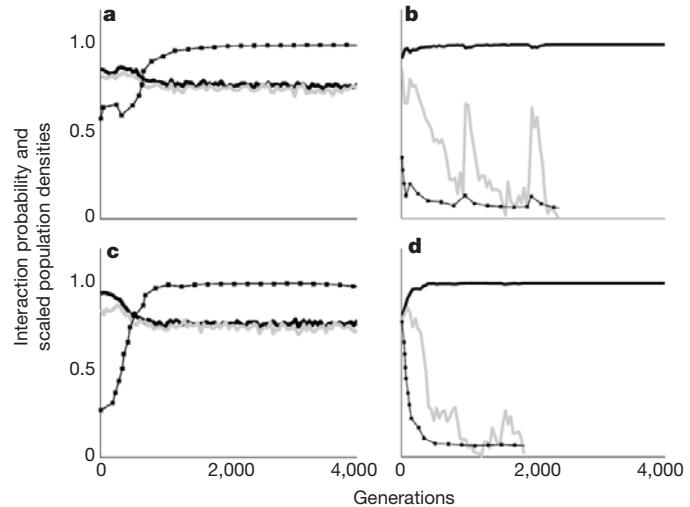


Figure 2 | Interaction probabilities and population densities in representative numerical simulations. Beaded lines show the mean interaction probability in each generation. Black lines show the victim density, and grey lines show the exploiter density, with densities scaled to the carrying capacity of the victim. **a, b**, The difference trait case for $n = 1$ (**a**; $\delta_{d,\max} = -0.012$) and $n = 4$ (**b**; $\delta_{d,\max} = 0.046$). **c, d**, The matching trait case for $n = 1$ (**c**; $\delta_{m,\max} = -0.029$) and $n = 4$ (**d**; $\delta_{m,\max} = 0.069$). Maximum evolutionary escape rates were calculated in the first pictured generation of each simulation, but changed as G-matrices evolved.

species increasingly as the number of traits in the system increases. In the difference trait case, this happens because antagonistic interspecific interactions cause negative correlations to evolve between the victim's defensive traits. These negative correlations constrain victim evolution. As the number of traits in the system increases, these negative correlations dissipate, and the constraint on victim evolution is removed (Methods and Supplementary Table 5). In the matching trait case, antagonistic interactions cause increased variance in the traits of the victim population and decreased variance in the traits of the exploiter population^{10,26}. Greater genetic variance in the victim population allows the victim to evolve more rapidly relative to the exploiter. As the number of traits in the system increases, the genetic variance in the victim population increases relative to that in the exploiter population (Methods and Supplementary Table 6). As a result, the victim can evolve away from the exploiter more rapidly when the number of traits in the system is large.

In nature, victim and exploiter population sizes can be sensitive to interaction rates, and therefore coevolution can affect population dynamics. To test whether our qualitative results hold in such cases, we performed additional simulations in which the absolute fitnesses of individuals in each population depended on the number of interactions in which those individuals participated (Methods). These simulations showed two qualitatively different ecological outcomes (Fig. 2). In the first, the exploiter evolves faster than the victim in all traits. The interaction probability approaches one, and the exploiter wins the evolutionary contest (Fig. 2a, c). In the second, the victim evolves faster than the exploiter in at least one trait. The interaction probability declines and the exploiter population shrinks. As exploiters become less common, selection for increased victim defence is reduced, and the

Table 2 | Outcome of coevolution in the matching trait case

n	$r_p = 0.00$	$r_p = 0.08$	$r_p = 0.26$	$r_p = 1.00$
1	0.7447* (22*)	– (–)	– (–)	– (–)
2	0.6700 (22)	0.6602 (25)	0.5724 (37)	0.2980 (59)
4	0.5262 (26)	0.4633 (33)	0.2373 (76)	0.0583 (92)
8	0.3239 (43)	0.2425 (55)	0.0399 (98)	0.0074 (98)

We conducted 100 simulations for each combination of trait number n and correlation strength r_p . We recorded a victim win if the mean interaction probability decreased over 2,000 generations of interaction. Results are shown as mean final interaction probability (percentage victim wins).

*Correlation is not meaningful when $n = 1$.

rate of victim evolution slows. As a result, the exploiter may be able to persist with a low interaction probability and a low population density (Fig. 2b, d). If the populations are sufficiently small, the chance fixation of favourable mutations in the victim or unfavourable mutations in the exploiter can reduce the interaction probability until the exploiter cannot achieve a positive population growth rate. If this happens, the exploiter population crashes and there is little opportunity for new favourable mutations to rescue the rapidly shrinking exploiter population from extinction. In the difference trait, matching trait and mixed trait cases, the probability that the victim evolves to lower its interaction probability with its exploiter, and the probability that the exploiter goes extinct, increase as the number of traits or the strength of the correlations between traits increases (Methods and Supplementary Tables 2–4). This supports the results of our fixed-population analyses.

Our study is motivated by host–parasite systems, but our models are general to any victim–exploiter species pair. For example, in nature predators must first detect, then capture, and finally subdue their victims, and victims have defensive traits to thwart their predators at each stage of the attack¹³. The presence of multiple defence mechanisms may help prey species to evolve and maintain low interaction rates with their predators. Similarly, herbivores often experience meaningful loss of fitness as a result of the various defence mechanisms of their host plants^{27,28}. Coevolution in highly dimensional trait space may help plants to limit the damage they receive from herbivores, and so may help to explain why the world is green²⁹.

METHODS SUMMARY

We modelled a victim–exploiter system in which selection is due only to interspecific interactions. Exploiters encounter potential victims at random, and they interact with a probability that depends on a suite of n genetically determined traits in each individual.

Quantitative genetic analysis. Under appropriate assumptions (Methods), the change in population mean trait values during one generation is^{20,21}

$$\begin{aligned}\Delta \bar{z}_e &= \mathbf{G}_e \boldsymbol{\beta}_e \\ \Delta \bar{z}_v &= \mathbf{G}_v \boldsymbol{\beta}_v\end{aligned}\quad (3)$$

where each $\boldsymbol{\beta}_x$ is a vector of selection gradients that depends on the interaction probability function, the fitness effect of interactions and the trait values in each population. For difference trait systems, we determined whether the victim wins by analysing the conditions under which $\Delta \bar{z}_{vi} > \Delta \bar{z}_{ei}$ for at least one i . For matching trait systems, (3) has an equilibrium at $\bar{z}_e = \bar{z}_v$. We determined whether the victim wins by analysing the conditions under which the maximum eigenvalue of the Jacobian of (3) evaluated at $\bar{z}_e = \bar{z}_v$ is greater than zero.

Numerical simulations. We complemented our analytical approach with genetically explicit individual-based simulations^{21,22}. Trait values in either species were governed by a set of unlinked, diallelic loci with pleiotropic alleles. Pleiotropy and linkage disequilibria caused correlations between trait values in each population. We considered two well-studied heuristic interaction probability functions. An exploiter with trait values $(z_{e1}, z_{e2}, \dots, z_{en})$ interacts with a victim with trait values $(z_{v1}, z_{v2}, \dots, z_{vn})$ with probability

$$P(I|z_{e1}, z_{e2}, \dots, z_{en}, z_{v1}, z_{v2}, \dots, z_{vn}) = \prod_{i=1}^n (1 + \exp(-g_i(z_{ei} - z_{vi})))^{-1} \quad (4)$$

in the difference trait case and

$$P(I|z_{e1}, z_{e2}, \dots, z_{en}, z_{v1}, z_{v2}, \dots, z_{vn}) = \prod_{i=1}^n \exp(-\gamma_i(z_{ei} - z_{vi})^2) \quad (5)$$

in the matching trait case. Parameters g_i and γ_i determine the sensitivity of the interaction probability to trait i . Reproduction is sexual, and individuals produce offspring in proportion to their fitnesses after interactions. This framework allows trait means, variances and covariances to evolve. We iterated generations and recorded whether the interaction probability between exploiters and victims increased (that is, the exploiter won) or decreased (that is, the victim won) over time.

Received 3 October 2011; accepted 12 January 2012.

Published online 4 March 2012.

1. Hafner, M. S. *et al.* Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265**, 1087–1090 (1994).

2. Page, R. D. M., Lee, P. L. M., Becher, S. A., Griffiths, R. & Clayton, D. H. A different tempo of mitochondrial DNA evolution in birds and their parasitic lice. *Mol. Phylogenet. Evol.* **9**, 276–293 (1998).
3. Dowton, M. & Austin, A. D. Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the hymenoptera. *J. Mol. Evol.* **41**, 958–965 (1995).
4. King, K. C., Jokela, J. & Lively, C. M. Trematode parasites infect or die in snail hosts. *Biol. Lett.* **7**, 265–268 (2011).
5. Hamilton, W. D., Axelrod, R. & Tanese, R. Sexual reproduction as an adaptation to resist parasites (a review). *Proc. Natl. Acad. Sci. USA* **87**, 3566–3573 (1990).
6. Thompson, J. N. Constraints on arms races in coevolution. *Trends Ecol. Evol.* **1**, 105–107 (1986).
7. Costamagna, A. C. & Landis, D. A. Predators exert top-down control of soybean aphid across a gradient of agricultural management systems. *Ecol. Appl.* **16**, 1619–1628 (2006).
8. Gomez, J. M. & Zamora, R. Top-down effects in a tritrophic system—parasitoids enhance plant fitness. *Ecology* **75**, 1023–1030 (1994).
9. Abrams, P. A. The evolution of predator–prey interactions: theory and evidence. *Annu. Rev. Ecol. Syst.* **31**, 79–105 (2000).
10. Frank, S. A. Evolution of host–parasite diversity. *Evolution* **47**, 1721–1732 (1993).
11. Hochberg, M. E. Hide or fight? The competitive evolution of concealment and encapsulation in parasitoid–host associations. *Oikos* **80**, 342–352 (1997).
12. Agrawal, A. F. & Lively, C. M. Modelling infection as a two step process combining gene-for-gene and matching allele genetics. *Proc. R. Soc. Lond. B* **270**, 323–334 (2003).
13. Vermeij, G. J. Unsuccessful predation and evolution. *Am. Nat.* **120**, 701–720 (1982).
14. Berenbaum, M. R., Zangerl, A. R. & Nitao, J. K. Constraints on chemical coevolution—wild parsnips and the parsnip webworm. *Evolution* **40**, 1215–1228 (1986).
15. Jones, S. R. M. The occurrence and mechanisms of innate immunity against parasites in fish. *Dev. Comp. Immunol.* **25**, 841–852 (2001).
16. Doebeli, M. & Ispolatov, I. Complexity and diversity. *Science* **328**, 494–497 (2010).
17. Toju, H. *et al.* Climatic gradients of arms race coevolution. *Am. Nat.* **177**, 562–573 (2011).
18. Nuismier, S. L. & Ridenhour, B. J. The contribution of parasitism to selection on floral traits in *Heuchera grossularifolia*. *J. Evol. Biol.* **21**, 958–965 (2008).
19. Rasmann, S., Erwin, A. C., Halitschke, R. & Agrawal, A. A. Direct and indirect root defences of milkweed (*Asclepias syriaca*): trophic cascades, trade-offs and novel methods for studying subterranean herbivory. *J. Ecol.* **99**, 16–25 (2011).
20. Lande, R. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**, 402–416 (1979).
21. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
22. Kirkpatrick, M. Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136**, 271–284 (2009).
23. Arnold, S. J., Burger, R., Hohenlohe, P. A., Ajie, B. C. & Jones, A. G. Understanding the evolution and stability of the G-matrix. *Evolution* **62**, 2451–2461 (2008).
24. Jones, A. G., Arnold, S. J. & Burger, R. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747–1760 (2003).
25. Jones, A. G., Arnold, S. J. & Burger, R. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* **58**, 1639–1654 (2004).
26. Nuismier, S. L., Doebeli, M. & Browning, D. The coevolutionary dynamics of antagonistic interactions mediated by quantitative traits with evolving variances. *Evolution* **59**, 2073–2082 (2005).
27. Zangerl, A. R. & Berenbaum, M. R. Phenotype matching in wild parsnip and parsnip webworms: causes and consequences. *Evolution* **57**, 806–815 (2003).
28. Zalucki, M. P., Brower, L. P. & Alonso-Mejia, A. Detrimental effects of latex and cardiac glycosides on survival and growth of first-instar monarch butterfly larvae *Danaus plexippus* feeding on the sandhill milkweed *Asclepias humistrata*. *Ecol. Entomol.* **26**, 212–224 (2001).
29. Hairston, N. G., Smith, F. E. & Slobodkin, L. B. Community structure, population control, and competition. *Am. Nat.* **94**, 421–425 (1960).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. R. Ives and F. Débarre for comments. R.T.G. and D.C.J. are postdoctoral fellows at the National Institute for Mathematical and Biological Synthesis (NIMBioS). NIMBioS is sponsored by the National Science Foundation (NSF), the US Department of Homeland Security and the US Department of Agriculture through NSF award no. EF-0832858, with support from The University of Tennessee, Knoxville. Additional support for this research was provided by NSF grants DMS 0540392 and DEB 1118947 to S.L.N.

Author Contributions R.T.G. and S.L.N. conceived the study and derived the analytical results. R.T.G. coded and ran numerical simulations and analysed output data. R.T.G. and D.C.J. derived proofs 1–3. R.T.G. and S.L.N. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.T.G. (rtgilman@nimbios.org).

Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills

Aaron C. Koralek^{1*}, Xin Jin^{5*}, John D. Long II¹, Rui M. Costa^{5,6} & Jose M. Carmena^{1,2,3,4}

The ability to learn new skills and perfect them with practice applies not only to physical skills but also to abstract skills¹, like motor planning or neuroprosthetic actions. Although plasticity in corticostriatal circuits has been implicated in learning physical skills^{2–4}, it remains unclear if similar circuits or processes are required for abstract skill learning. Here we use a novel behavioural task in rodents to investigate the role of corticostriatal plasticity in abstract skill learning. Rodents learned to control the pitch of an auditory cursor to reach one of two targets by modulating activity in primary motor cortex irrespective of physical movement. Degradation of the relation between action and outcome, as well as sensory-specific devaluation and omission tests, demonstrate that these learned neuroprosthetic actions are intentional and goal-directed, rather than habitual. Striatal neurons change their activity with learning, with more neurons modulating their activity in relation to target-reaching as learning progresses. Concomitantly, strong relations between the activity of neurons in motor cortex and the striatum emerge. Specific deletion of striatal NMDA receptors impairs the development of this corticostriatal plasticity, and disrupts the ability to learn neuroprosthetic skills. These results suggest that corticostriatal plasticity is necessary for abstract skill learning, and that neuroprosthetic movements capitalize on the neural circuitry involved in natural motor learning.

The ability to learn new actions and perfect them with practice allows us to master skills like playing the piano or riding a bicycle. Learning these skills usually implies moving faster, more accurately and less variably⁵. However, mastering other types of skills, like playing board games or controlling neuroprosthetic devices, often does not directly involve changes in physical movement^{1,6}. Cortico-basal ganglia circuits have been implicated in the learning, selection and execution of physical skills^{2–4,7,8}. In particular, plasticity in the motor cortices and the striatum, the major input region of the basal ganglia, has been shown to accompany the learning of physical skills^{2,9}. The motor cortex and frontal cortices have also been implicated in the learning of abstract skills^{10–13}, and in learning to control neuroprosthetic devices irrespective of physical movement^{14–17}. Some studies suggest that not only cortical areas, but also the striatum, are involved in learning abstract skills^{18–20}. However, it is still unclear if the striatum is required for abstract skill learning, and if corticostriatal circuits undergo plasticity during the learning of such skills as they do during the learning of physical skills. Here, we use a novel behavioural task in conjunction with electrophysiology and genetic manipulation in rodents to investigate the role of corticostriatal circuits and corticostriatal plasticity in the learning of intentional neuroprosthetic actions: that is, actions performed with disembodied actuators based on the modulation of specific neural activity and irrespective of physical movement⁶.

We developed a novel operant brain-machine interface task in which rodents were required to modulate activity in M1, rather than execute a physical movement, to obtain reward (Fig. 1a). Modulation

of M1 ensemble activity resulted in changes in the pitch of an auditory cursor, which provided constant auditory feedback to rodents about task performance. Reward was delivered when rodents precisely modulated M1 activity to move this auditory cursor to one of two

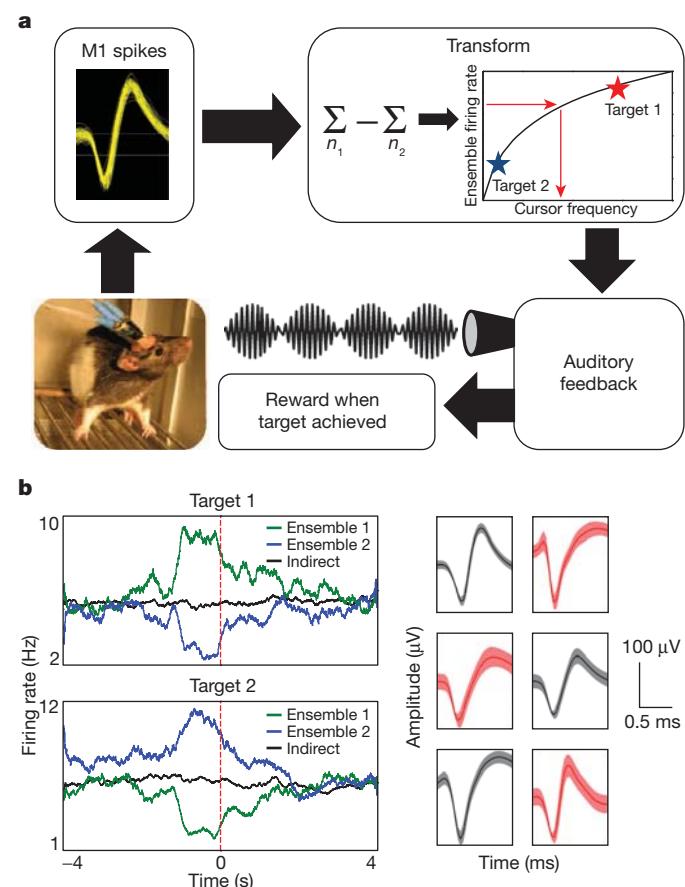


Figure 1 | Volitional modulation of M1 neural activity in awake behaving rodents. a, Task schematic. M1 unit activity was entered into an online transform algorithm that related ensemble activity to the pitch of an auditory cursor. Two opposing ensembles were chosen, with activity of one ensemble increasing the cursor pitch and activity of the other ensemble decreasing the cursor pitch. Constant auditory feedback about cursor location was supplied to rodents, and distinct rewards were supplied when rodents brought M1 activity into one of two target states. b, Mean M1 ensemble firing rates for units in ensemble 1 (green), ensemble 2 (blue) and M1 units not used in the transform (black) in relation to the achievement of target 1 (top) or target 2 (bottom). Representative waveforms recorded from M1 in rats are shown on the right, with shaded regions denoting the standard deviation.

¹Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. ²Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA. ³Program in Cognitive Science, University of California, Berkeley, California 94720, USA. ⁴JC Berkeley and UC San Francisco Joint Graduate Group in Bioengineering, University of California, Berkeley, California 94720, USA. ⁵Laboratory for Integrative Neuroscience, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5625 Fishers Lane, Bethesda, Maryland 20892-9412, USA. ⁶Champalimaud Neuroscience Programme, Champalimaud Center for the Unknown, Avenida de Brasília, 1400-038 Lisbon, Portugal.

*These authors contributed equally to this work.

target tones, and a trial was marked incorrect if no target had been hit within a set time limit (30 s). One of these targets was associated with a reward of sucrose solution, whereas the other target was associated with a pellet reward (see Methods). Two neural ensembles consisting of two- to four-well-isolated units each were used to control the auditory cursor (Supplementary Figs 1 and 2). The action of these two ensembles opposed each other, such that increased activity in one ensemble produced increases in cursor pitch, whereas increased activity in the other ensemble caused decreases in cursor pitch. Thus, to achieve a high-pitched target, rodents had to increase activity in the first ensemble and decrease activity in the second; the opposite was required to hit a low-pitched target (Fig. 1b and Supplementary Fig. 3). These firing-rate modulations had to be maintained for several time bins (200 ms bin size) for a target to be hit (Supplementary Methods). Hence, in this operant task, rodents had to bring the two M1 ensembles into a desired state irrespective of motor output.

We trained six male Long-Evans rats on the task, and verified that they exhibited marked improvement in the percentage of correct trials over the course of 11 days (Fig. 2a). As typically observed in motor skill learning²¹, there was a phase of rapid improvement followed by a phase of slower learning, representing early (days 2–4) and late (days 8–11) phases of learning. The percentage of correct trials increased significantly from early to late in learning (Fig. 2b; $P < 0.001$), resulting in performance well above chance (Fig. 2c; $P < 0.001$; see Supplementary Methods), whereas mean time-to-target decreased (Supplementary Fig. 4). Analyses of M1 firing rates further showed that rats were producing the desired neuronal ensemble rate modulations during task performance (Fig. 1b and Supplementary Fig. 3). Furthermore, sensory feedback was found to be critical for animals to learn this task because when rats were not given auditory feedback during training (although they would still get a reward if they would modulate neural activity correctly), the percentage of correct trials did not increase over the course of 9 days of training (Supplementary Fig. 5).

We next investigated if animals were performing physical movements that would modulate the activity of those particular M1 ensembles. First, we monitored overall rodent movement with an accelerometer mounted on the recording headstage, which allowed us to measure if the animals produced any body or head movement during target achievement²². Accelerometer traces exhibited no changes before and during target reaching, but did show prominent deflections after target reaching as the animals retrieved the reward (Fig. 2d), demonstrating that rodents were not relying on gross motor behaviour to perform the task. We also monitored movements of the vibrissae with electromyographic (EMG) recordings of the mystacial pad (electrodes targeted M1 areas controlling vibrissae movement; Supplementary Fig. 2 and Supplementary Methods), and observed no significant EMG signals before target achievement, although there were clear EMG signals afterwards as animals retrieved and consumed the reward (Fig. 2e and Supplementary Fig. 6b). Importantly, there was no correlation between EMG activity and the spiking of the M1 neurons controlling the auditory cursor: the correlation coefficient for all trials in a behavioural session was 0.092 ± 0.003 (mean \pm s.e.m.), and the distribution of correlation coefficients across a session was not significantly different from zero (Supplementary Fig. 6a; $P = 0.57$). This was observed across all training days, including during early learning (Supplementary Fig. 7). These data suggest that rats do not rely on physical movements to learn the task, although it is difficult to exclude the possibility that animals use some movement to generate neural activity to drive the auditory cursor during exploratory phases of the task in early learning. Nonetheless, the data show that animals eventually learn to perform the task in the absence of overt movement. To demonstrate further that rats did not require vibrissae movements to control M1 activity, we injected lidocaine into the whisker pad to inactivate sensory and motor nerve endings locally during a session in late learning (see Supplementary Methods). There was no significant change in performance during the temporary inactivation (Fig. 2f).

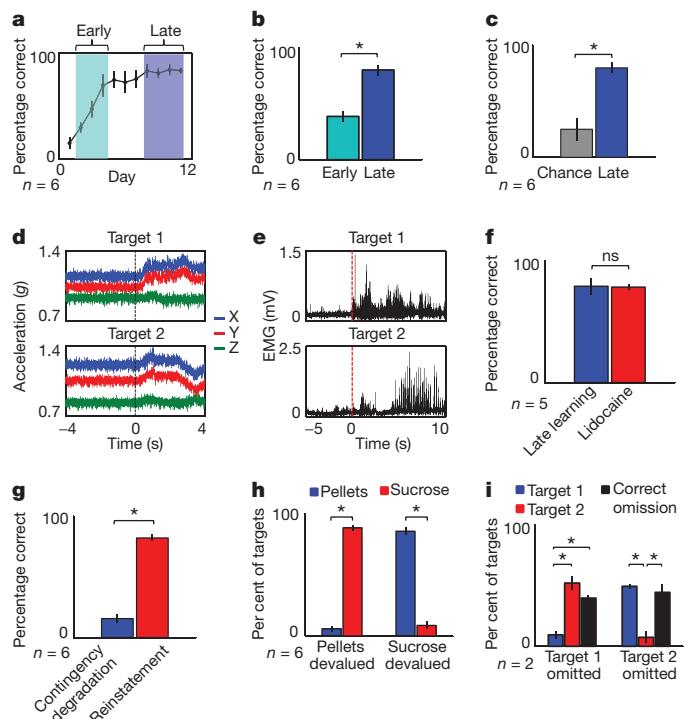


Figure 2 | Learning intentional neuroprosthetic actions independently of movement. **a**, Mean percentage of correct responses for all rats across days 1–11 of learning. Shaded regions denote the range of days from which the early and late learning analyses were performed. **b**, The percentage of correct responses for all rats increased significantly from early (light blue) to late (dark blue) in learning. **c**, Percentage of correct responses in late learning is significantly greater than expected by chance. **d**, Representative accelerometer traces show no gross motor behaviour leading to target achievement (time zero), but clear deflections as animals initiate movement to retrieve reward. **e**, Representative EMG traces show no muscle activity in the mystacial pad before target achievement, but clear deflections as animals retrieve and consume reward. **f**, Mean performance in all rats when lidocaine was injected into the whisker pad before a behavioural session late in learning (red) compared with performance during a no-lidocaine session (dark blue). **g**, Significant reduction in response rate when the causal relation between target achievement and reward delivery was degraded (dark blue). When contingency was reinstated, performance rapidly returned to pre-degradation levels (red). **h**, Percentage of total correct trials that were directed at the target associated with pellet reward (blue) or sucrose solution reward (red) during choice sessions where rats had free access to pellets (left; ‘Pellets devalued’), or to sucrose before the session (right; ‘Sucrose devalued’). **i**, Percentage of total trials that involved responses towards target 1 (blue), target 2 (red) or response omissions (black) when omission tests were performed for target 1 (left) or target 2 (right).

$P > 0.9$), with rats achieving $78.1 \pm 2.2\%$ correct with lidocaine (mean \pm s.e.m.) versus $78.8 \pm 6.5\%$ without lidocaine. Taken together, these data indicated that rodents were able to learn to control M1 activity operantly irrespective of any overt movement.

Goal-directed actions are sensitive to changes in the relation between performing the action and obtaining a reward (contingency) and to changes in the expected value of the reward, whereas habits are not^{23,24}. We asked if these neuroprosthetic actions were performed intentionally, because the animal volitionally controlled M1 activity to get the outcome (goal-directed), or habitually owing to the reinforcement history. To test this, we first degraded the contingency between executing the action and obtaining the outcome: that is, the auditory cursor was still under control of M1 ensemble activity, but the probability of obtaining reward was similar irrespective of target achievement, which had no effect on the rate of reward. After 2 days of contingency degradation, rats markedly diminished their responding and the percentage of correct trials decreased significantly (Fig. 2g;

$P < 0.001$). When contingency was reinstated, rats resumed responding and the percentage of correct trials returned to plateau levels seen in late learning (Fig. 2g).

To investigate further the intentional nature of these neuroprosthetic skills, we performed a test where each of the outcomes was devalued using sensory-specific satiety. Rats were given free access either to sucrose solution or pellets for 1 h before the behavioural session, thereby reducing the expected value of that outcome²⁵. After specific devaluation of each outcome/reward, rats chose the target leading to that reward much less than the target leading to the reward that was not devalued (Fig. 2h; $P < 0.001$), indicating that their actions were sensitive to changes in outcome value. Importantly, there were no significant differences in reward preference during normal task performance when neither of the outcomes was devalued (Supplementary Fig. 8; $P > 0.25$). Finally, we asked whether rats were able to intentionally inhibit the reaching of one of the two targets to obtain the specific reward associated with that target. To examine this we performed an omission test, where the reward previously associated with reaching a particular target was only delivered when rats successfully inhibited reaching that target throughout the duration of the trial. If the target was reached during the 30 s of trial duration, no reward was delivered and a new trial was initiated. Importantly, reaching the other target continued to lead to reward as during training. Animals behaved in a goal-directed manner in the omission test for both targets, because they reduced the number of reaches for the target they had to omit versus the no-omission target, while increasing the number of correctly omitted responses (Fig. 2i; $P < 0.001$ for both comparisons). Taken together, these data show that the neuroprosthetic actions in our task are sensitive to changes in the causal relation between performing the action and obtaining the reward (contingency degradation and omission test), and to changes in the expected value of the outcome (sensory-specific devaluation), indicating that they are intentional and goal-directed rather than habitual.

We next examined if learning to operantly control M1 activity irrespective of overt movement involves striatal plasticity, akin to what is observed for natural motor learning^{2–4,7,26–28}. We verified that the improvement in behavioural performance seen across learning was accompanied by a significant increase in firing rates in the dorsal striatum (DS) in late learning compared with early learning (Fig. 3a; $P < 0.001$). In addition to this general increase in firing rates, we noticed that firing rates of DS neurons exhibited greatest modulation during target reaching compared with baseline control periods (Fig. 3b), as observed during natural motor learning²⁶. This modulation was significantly greater in late learning than early learning (Fig. 3b, c; $P < 0.05$), indicating that DS neurons changed their activity during the volitional control of M1 activity, and that this change increased with learning.

We next investigated if learning, and the observed changes in DS target-related activity, were accompanied by corticostriatal plasticity, that is, changes in the functional interactions between M1 and DS neurons. We noticed that cross-correlation histograms between the two regions in late learning exhibited pronounced oscillatory spike coupling (Fig. 3d). To quantify this interaction, we calculated the coherence between spiking activity in the two regions in both early and late learning (Supplementary Methods). The resulting coherograms exhibited a clear increase in coherence at low-frequency bands in late learning relative to early learning (Fig. 3e), and these frequencies corresponded to the oscillatory frequency seen in the cross-correlograms (Supplementary Methods). Furthermore, mean coherence in the theta band (4–8 Hz) was significantly greater in late learning than early learning (Fig. 3f; $P < 0.001$). This increase in coherence appeared to be related to learning to perform the task rather than higher reward expectation or proportion of correct trials in late learning, because coherence values remained high surrounding target achievement during the contingency degradation manipulation, where reward delivery was not contingent upon target achievement (Supplementary

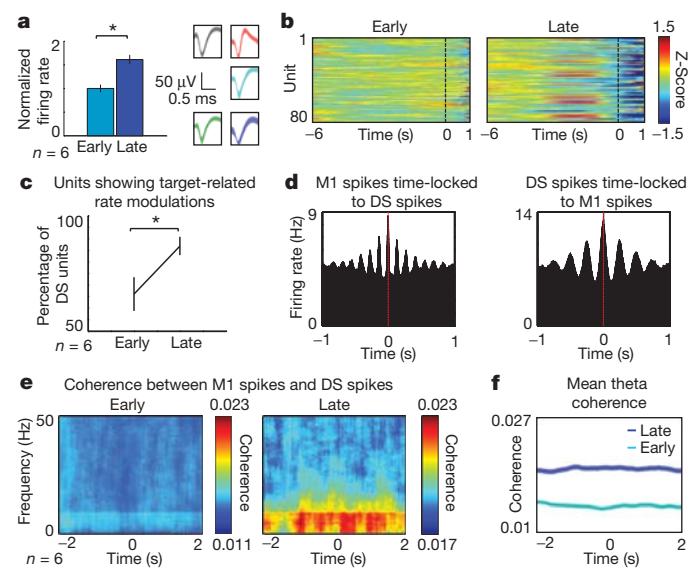


Figure 3 | Learning abstract skills is accompanied by corticostriatal plasticity. **a**, Mean normalized firing rates in DS increased significantly from early (light blue) to late (dark blue) learning. Representative waveforms recorded from the DS are shown on the right (shaded regions denote standard deviation). **b**, Z-scored firing rates for individual DS units in relation to target achievement (time zero) showing marked modulation before target achievement in late learning. **c**, The percentage of DS units exhibiting target-related firing-rate modulation increased significantly with learning. **d**, Cross-correlation histograms in late learning for M1 spiking activity in relation to DS spikes (left), and DS spiking activity in relation to M1 spikes (right), showing oscillatory coupling between the two regions. **e**, Coherence between M1 spikes and DS spikes in early (left) and late (right) learning shows a clear increase in low-frequency coherence from early to late learning. **f**, Significant increase in mean coherence in the theta range in late (dark blue) versus early (light blue) learning. Shaded regions denote s.e.m.

Fig. 9; not significantly less than non-degraded trials, $P > 0.05$). In addition, coherence levels remained high during task performance in incorrect trials (data not shown), further suggesting that the increase in coherence observed is due to learning to perform the skill rather than outcome anticipation. Thus, neuroprosthetic skill learning is accompanied by dynamic changes in functional interactions between M1 and the DS neurons, suggesting an important role for corticostriatal plasticity in this novel task.

We therefore investigated if corticostriatal plasticity would be necessary for neuroprosthetic skill learning. NMDARs (*N*-methyl-D-aspartic acid receptors) in striatal medium spiny neurons are critical for corticostriatal long-term potentiation²⁹. We used a knockin line that expresses Cre recombinase in both striatonigral and striatopallidal medium spiny neurons (*RGS9L-cre*), but not in all striatal neurons (for example, absent from parvalbumin interneurons; Supplementary Fig. 10), and crossed it with mice carrying a floxed allele of the *NMDAR1* gene³⁰. The resulting mice lack NMDA currents in most projection neurons³⁰ (but not all striatal cells, hence we refer to them as *RGS9L-Cre/Nr1^{ff}*, not as striatal NR1 knockouts; see Supplementary Methods), and have impaired corticostriatal long-term potentiation³⁰. As previously described, these animals do not display any major motor deficits (Supplementary Videos 1 and 2) and can learn to perform rapid sequential movements (Supplementary Fig. 11), albeit being unable to learn precise motor sequences²⁷. We investigated neuroprosthetic skill learning in *RGS9L-Cre/Nr1^{ff}* mice and littermate controls. Although control mice showed performance improvement across learning irrespective of physical movement as observed for rats ($P < 0.001$; Fig. 4a, b), *RGS9L-Cre/Nr1^{ff}* mice exhibited marked learning deficits on the task, with no significant increase in the percentage of correct trials from early to late learning (Fig. 4a; $P = 0.98$). Furthermore, acute pharmacological blockade of NMDARs in trained

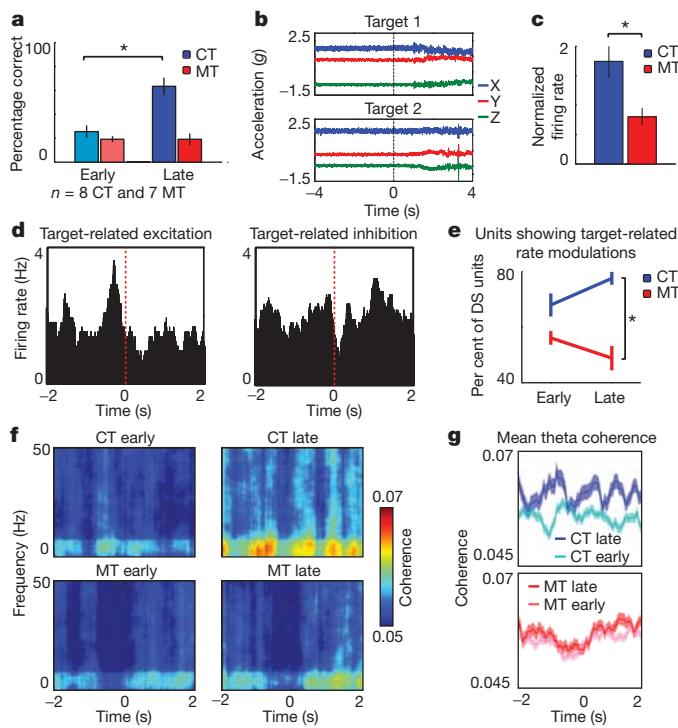


Figure 4 | Selective deletion of NMDARs in the striatum impairs brain-machine interface learning. **a**, *RGS9L-Cre/Nr1^{ff}* mice (red) exhibit no significant increase in the percentage of correct trials over the course of learning, despite clear performance improvement in littermate controls (blue). CT, controls; MT, mutants. **b**, Accelerometer traces from control mice showing no clear oscillation before target achievement, but clear deflections as mice retrieve reward. **c**, Late firing rate normalized to early firing rate in controls and mutants. There is no significant increase in DS firing rates in *RGS9L-Cre/Nr1^{ff}* mice (red) from early to late learning, although DS firing rates increase markedly in control mice (blue). **d**, DS units of controls exhibit strong target-related firing-rate modulations, including both excitation (left) and inhibition (right). **e**, The percentage of DS units showing significant target-related firing-rate modulations increases significantly across learning in control mice (blue), but not in *RGS9L-Cre/Nr1^{ff}* mutants (red). **f**, Coherograms showing coherence between M1 spikes and DS spikes in early (left) and late (right) learning both for control mice (top) or *RGS9L-Cre/Nr1^{ff}* mice (bottom). Coherence in low-frequency bands increases from early to late learning in control mice, but not in *RGS9L-Cre/Nr1^{ff}* mice. **g**, Mean coherence in the theta range for control (top) and mutant (bottom) mice. There is a significant increase in coherence from early (light blue) to late (dark blue) learning in control mice but not in mutant mice (early learning, light red; late learning, dark red).

control animals did not affect performance of the neuroprosthetic skill (even at relatively high doses that affect striatal burst firing; Supplementary Figs 12 and 13), suggesting that the deficits observed in *RGS9L-Cre/Nr1^{ff}* mice are not due to inability to perform the skill but rather to the inability to learn the task.

Consistent with the findings above in rats (Fig. 3a), DS neurons in littermate control mice exhibited a significant increase in firing rate across learning, whereas in mutants they did not (Fig. 4c; main effect of genotype $F_{1,10} = 32.45$, $P < 0.001$; early versus late $P < 0.05$ for controls (CT) and $P = 0.23$ for mutants (MT)). Also, in control mice, the proportion of DS neurons with significant target-related firing-rate modulation increased with learning (Fig. 4d, e; $P < 0.05$), but this was not observed in *RGS9L-Cre/Nr1^{ff}* mice (Fig. 4e; $P = 0.28$). Finally, the development of functional corticostriatal interactions during learning was also abolished in *RGS9L-Cre/Nr1^{ff}* mutants, with no significant increase in coherence between M1 and DS spikes with learning (Fig. 4f, g, $F_{80,10} = 0.65$, $P = 0.44$), although littermate controls showed a clear increase as seen in rats (Fig. 4f, g; $F_{80,10} = 4.86$, $P < 0.05$). Taken together, these results demonstrate that the striking corticostriatal plasticity observed in rats during learning also occurs in control mice,

but this plasticity is absent in mice with a decrease in functional NMDARs in striatum. These mutant mice do not show improvement with training, therefore indicating that corticostriatal plasticity may be necessary for learning to modulate M1 states intentionally to obtain specific outcomes.

In summary, we used a novel operant task in rodents to demonstrate that corticostriatal networks exhibit profound plasticity during the learning of intentional neuroprosthetic skills and, further, that disrupting this plasticity impairs learning. This adds great support to the claims that cortico-basal ganglia circuits play a role in abstract cognitive processes^{18–20}. We observed that DS neurons strongly modulated their activity in relation to M1 activity, even when the latter was dissociated from physical movements, suggesting that the striatum is important for learning and selecting abstract actions that are controlled by cortical output. Hence, these data suggest that cortico-basal ganglia circuits may be involved in learning mental actions and skills that do not require physical movement, indicating that they may have a broader function involved in intention and decision-making than previously acknowledged.

Our results also have important implications for the field of brain-machine interfaces⁶. The abstract actions investigated here form the basis for skilful neuroprosthetic control¹⁶ and, as we have shown here, they recruit elements of the natural motor system outside of M1. Thus, our results suggest that neuroprosthetic movements capitalize on the neural circuitry for motor learning and therefore have great potential to feel naturalistic, generalize well to novel movements and environments, and benefit from our nervous system's highly developed storage and retrieval mechanisms for skilled behaviour.

METHODS SUMMARY

All experiments were done in accordance with the Animal Care and Use Committee at the University of California, Berkeley, and at the National Institute on Alcohol Abuse and Alcoholism, and according to National Institutes of Health guidelines. Six male Long-Evans rats, seven *RGS9L-Cre/Nr1^{ff}* mice and eight littermate controls were chronically implanted with tungsten microelectrode arrays in both M1 and the DS ipsilaterally. Two ensembles of two to four well-isolated M1 units each were chosen and, by modulating activity in these ensembles, rodents controlled the pitch of an auditory cursor, with increased activity in the first ensemble producing increases in the cursor pitch and increased activity in the other ensemble producing decreases in the cursor pitch. The rodents had to modulate these ensembles precisely to move the cursor to one of two target pitches to get reward (one associated with 20% sucrose and another with pellets). Rodents were free to choose either reward at any time. A trial was correct if a target was achieved within 30 s and incorrect trials were followed by a time-out. M1 activity levels had to return to baseline levels for a new trial to begin. After performance had reached plateau levels, the action-outcome contingency was degraded by providing outcomes on a variable time schedule to match the probabilities between getting a reward after target achievement and no target achievement. For outcome devaluation, rodents were given a sensory-specific satiation test where they received free access to one of the rewards (but not the other) for 1 h before a behavioural session. For the omission test, rodents stopped being rewarded for reaching one of the targets and the reward associated with that target was instead supplied when rodents successfully inhibited responses for the duration of the trial. See Supplementary Information for further details.

Received 21 April 2011; accepted 9 January 2012.

Published online 4 March 2012.

1. VanLehn, K. Cognitive skill acquisition. *Annu. Rev. Neurosci.* **47**, 513–539 (1996).
2. Yin, H. H. *et al.* Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature Neurosci.* **12**, 333–341 (2009).
3. Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z. & Graybiel, A. M. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* **437**, 1158–1161 (2005).
4. Kimchi, E. Y. & Laubach, M. Dynamic encoding of action selection by the medial striatum. *J. Neurosci.* **29**, 3148–3159 (2009).
5. Brashers-Krug, T., Shadmehr, R. & Bizzzi, E. Consolidation in human motor memory. *Nature* **382**, 252–255 (1996).
6. Fetz, E. E. Volitional control of neural activity: implications for brain-computer interfaces. *J. Physiol. (Lond.)* **579**, 571–579 (2007).
7. Hikosaka, O. *et al.* Parallel neural networks for learning sequential procedures. *Trends Neurosci.* **22**, 464–471 (1999).

8. Brasted, P. J. & Wise, S. P. Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *Eur. J. Neurosci.* **19**, 721–740 (2004).
9. Rioult-Pedotti, M. S., Friedman, D. & Donghue, J. P. Learning-induced LTP in neocortex. *Science* **290**, 533–536 (2000).
10. Georgopoulos, A. P., Taira, M. & Lukashin, A. Cognitive neurophysiology of the motor cortex. *Science* **260**, 47–52 (1993).
11. Gandolfo, F., Li, C., Benda, B. J., Schioppa, C. P. & Bizzi, E. Cortical correlates of learning in monkeys adapting to a new dynamical environment. *Proc. Natl Acad. Sci. USA* **97**, 2259–2263 (2000).
12. Fincham, J. M. & Anderson, J. R. Distinct roles of the anterior cingulate and prefrontal cortex in the acquisition and performance of a cognitive skill. *Proc. Natl Acad. Sci. USA* **103**, 12941–12946 (2006).
13. Badre, D., Kayser, A. S. & D'Esposito, M. Frontal cortex and the discovery of abstract action rules. *Neuron* **66**, 315–326 (2010).
14. Taylor, D. M., Tillery, S. I. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**, 1829–1832 (2002).
15. Carmena, J. M. et al. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1**, 193–208 (2003).
16. Ganguly, K. & Carmena, J. M. Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biol.* **7**, e1000153 (2009).
17. Ganguly, K., Dimitrov, D. F., Wallis, J. D. & Carmena, J. M. Reversible large-scale modification of cortical networks during neuroprosthetic control. *Nature Neurosci.* **14**, 662–667 (2011).
18. Beauchamp, M. H., Dagher, A., Aston, J. A. & Doyon, J. Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. *Neuroimage* **20**, 1649–1660 (2003).
19. Poldrack, R. A., Prabhakaran, V., Seger, C. A. & Gabriel, J. D. Striatal activation during acquisition of a cognitive skill. *Neuropsychology* **13**, 564–574 (1999).
20. Pasupathy, A. & Miller, E. K. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* **433**, 873–876 (2005).
21. Karni, A. et al. The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proc. Natl Acad. Sci. USA* **95**, 861–868 (1998).
22. Venkatraman, S., Jin, X., Costa, R. M. & Carmena, J. M. Investigating neural correlates of behavior in freely behaving rodents using inertial sensors. *J. Neurophysiol.* **104**, 569–575 (2010).
23. Balleine, B. W. & Dickinson, A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407–419 (1998).
24. Yin, H. H., Knowlton, B. J. & Balleine, B. W. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. *Behav. Brain Res.* **166**, 189–196 (2006).
25. Hilario, M. R., Clouse, E., Yin, H. H. & Costa, R. M. Endocannabinoid signaling is critical for habit formation. *Front. Integr. Neurosci.* **1**, 1–12 (2007).
26. Costa, R. M., Cohen, D. & Nicolelis, M. A. Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr. Biol.* **14**, 1124–1134 (2004).
27. Jin, X. & Costa, R. M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466**, 457–462 (2010).
28. Miyachi, S., Hikosaka, O. & Lu, X. Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp. Brain Res.* **146**, 122–126 (2002).
29. Calabresi, P., Pisani, A., Mercuri, N. B. & Bernardi, G. Long-term potentiation in the striatum is unmasked by removing the voltage-dependent magnesium block of NMDA receptor channels. *Eur. J. Neurosci.* **4**, 929–935 (1992).
30. Dang, M. T. et al. Disrupted motor learning and long-term synaptic plasticity in mice lacking NMDAR1 in the striatum. *Proc. Natl Acad. Sci. USA* **103**, 15254–15259 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Venkatraman for the three-axis accelerometer, Y. Li for the *RGS9L-Cre* mice, K. Nakazawa for the *NMDAR1-*loxP** mice, G. Luo for genotyping, M. Davis for advice on staining and G. Martins for performing immunohistochemistry. This work was supported by National Science Foundation CAREER Award 0954243, the Multiscale Systems Research Center and the Defense Advanced Research Projects Agency contract N66001-10-C-2008 to J.M.C., and the Division of Intramural Clinical and Basic Research of the National Institute on Alcohol Abuse and Alcoholism, Marie Curie International Reintegration Grant 239527 and European Research Council STG 243393 to R.M.C.

Author Contributions A.C.K., X.J., J.D.L., R.M.C. and J.M.C. designed experiments. A.C.K., X.J. and J.D.L. conducted experiments. A.C.K., X.J., R.M.C. and J.M.C. analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.C. (carmena@eecs.berkeley.edu) or R.M.C. (ruicosta@fchampalimaud.org).

Phase transitions in the assembly of multivalent signalling proteins

Pilong Li^{1*}, Sudeep Banjade^{1*}, Hui-Chun Cheng^{1*}, Soyeon Kim¹, Baoyu Chen¹, Liang Guo², Marc Llaguno³, Javoris V. Hollingsworth⁴, David S. King⁵, Salman F. Banani¹, Paul S. Russo⁴, Qiu-Xing Jiang³, B. Tracy Nixon⁶ & Michael K. Rosen¹

Cells are organized on length scales ranging from ångström to micrometres. However, the mechanisms by which ångström-scale molecular properties are translated to micrometre-scale macroscopic properties are not well understood. Here we show that interactions between diverse synthetic, multivalent macromolecules (including multi-domain proteins and RNA) produce sharp liquid–liquid-demixing phase separations, generating micrometre-sized liquid droplets in aqueous solution. This macroscopic transition corresponds to a molecular transition between small complexes and large, dynamic supramolecular polymers. The concentrations needed for phase transition are directly related to the valency of the interacting species. In the case of the actin-regulatory protein called neural Wiskott–Aldrich syndrome protein (N-WASP) interacting with its established biological partners NCK and phosphorylated nephrin¹, the phase transition corresponds to a sharp increase in activity towards an actin nucleation factor, the Arp2/3 complex. The transition is governed by the degree of phosphorylation of nephrin, explaining how this property of the system can be controlled to regulatory effect by kinases. The widespread occurrence of multivalent systems suggests that phase transitions may be used to spatially organize and biochemically regulate information throughout biology.

Covalent and non-covalent interactions between multivalent small molecules are central elements of classical polymer chemistry and physics and of supramolecular chemistry^{2–4}. These fields have produced theories and experimental demonstrations of sharp transitions between small assemblies and macroscopic polymer gels (known as sol–gel transitions) as the degree of bonding increases. The transition point (critical point) depends on the physical properties of the monomeric species, including valency and affinity. The polymer can have a variety of physical forms, ranging from phase-separated liquid to crystalline solid. For non-covalent systems, phase separation can strongly influence the sol–gel transition by altering the degree of bonding^{5,6}. In biology, interactions between multivalent entities are found in diverse processes, including extracellular carbohydrate–lectin binding, intracellular signalling, RNA metabolism and chromatin organization in the nucleus^{7–10}. Biological multivalency has been studied most extensively in the context of extracellular ligands binding to cell surface receptors, where antibody–receptor¹¹ and carbohydrate–lectin⁷ systems can assemble into crosslinked networks. These networks are typically precipitates^{11,12}, but liquid-like gels have also been described¹³. Multivalency has been less studied in the context of intracellular molecules, which often share the characteristics of high valency, modest affinity, and long, flexible connections between binding elements¹⁴. Here we asked whether these systems also undergo sharp transitions to polymers, and if so, what the macroscopic properties of the polymers are and how such transitions could be regulated and affect function.

Initially, we examined interactions between the SRC homology 3 (SH3) domain and its proline-rich motif (PRM) ligand, two widely observed modules that often appear in tandem arrays in signalling proteins^{8,14}. We generated two classes of engineered proteins: one composed of repeats of a single SH3 domain ($SH3_m$, where $m = 1–5$), and the other composed of repeats of a PRM ligand (PRM_n , where $n = 1–5$) (dissociation constant (K_d) = 350 μM for the $SH3_1$ – PRM_1 interaction; Supplementary Fig. 1). Initially, we mixed $SH3_4$ with PRM_4 . At low concentrations, the solutions were clear; by contrast, at high concentrations, they were opalescent. Examination of these opalescent solutions using light microscopy showed the presence of numerous spherical droplets of approximately 1 μm to $>50 \mu\text{m}$ in diameter that had phase-separated from the bulk solution (Fig. 1a and Supplementary Fig. 2). Smaller droplets tended to coalesce into larger droplets over time, which is consistent with liquid-like properties (Fig. 1b). When the proteins were mixed in a 1:1 ratio, both the droplet and bulk phases contained equal amounts of each molecule, but the proteins were concentrated by about 100-fold in the droplets relative to

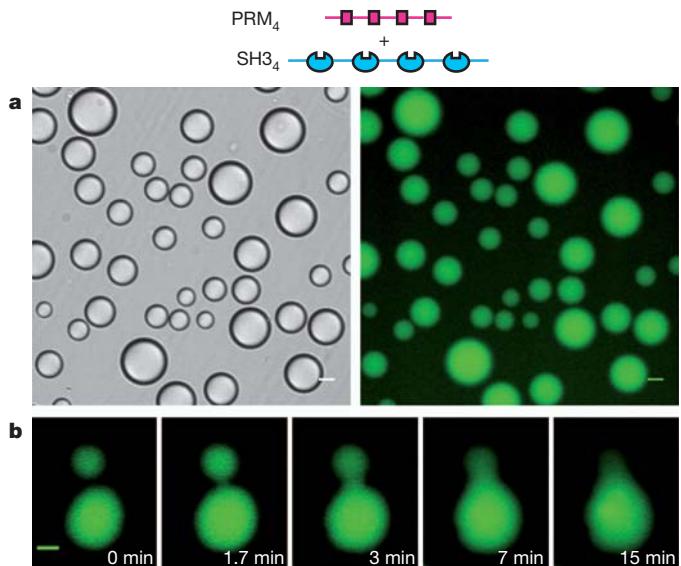


Figure 1 | Macroscopic and microscopic phase transitions in multivalent SH3–PRM systems. **a**, Liquid droplets observed by differential interference contrast microscopy (left) and wide-field fluorescence microscopy (right) when 300 μM $SH3_4$, 300 μM PRM_4 (both of which are module concentrations; molecule concentrations are 75 μM) and 0.5 μM OG– $SH3_4$ were mixed. Scale bars, 20 μm . **b**, Time-lapse imaging of merging droplets that were formed as in **a**. Scale bar, 10 μm .

¹Department of Biochemistry and Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-8812, USA. ²BioCAT of IIT at the Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. ³Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9148, USA.

⁴Department of Chemistry and Macromolecular Studies Group, Louisiana State University, Baton Rouge, Louisiana 70803, USA. ⁵Howard Hughes Medical Institute Mass Spectrometry Laboratory and Department of Molecular & Cell Biology, University of California, Berkeley, California 94720, USA. ⁶Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

*These authors contributed equally to this work.

the bulk phase (Fig. 1a, right; 116-fold for SH3_5 plus PRM_5 ; and 82-fold for SH3_5 plus an octameric dendrimer, $\text{PRM}(\text{N}-\text{WASP})_8$). Analogous droplets were also observed for an unrelated SH3_5 -ligand pair and for the tetravalent RNA binding protein PTB interacting with an RNA oligonucleotide (Supplementary Fig. 3). Thus, liquid–liquid demixing phase transitions may occur in many multivalent intracellular systems.

A large body of data indicates that the phase separation observed here is driven by the assembly of the multivalent proteins into large species, analogous to the behaviour observed in many small-molecule polymer systems^{2,5} as well as in covalent protein crosslinking¹⁵. First, the phase boundary is strongly dependent on the valency of the interacting species (Fig. 2a and Supplementary Fig. 4). This observation is consistent with theory and our simulations, which indicate that a higher valency allows the formation of larger species at a lower fractional saturation of the binding modules^{2,3,5,16} (Supplementary Information and Supplementary Figs 5–7).

Second, the phase transition can be blocked by a high-affinity monovalent ligand, $\text{PRM}(\text{H})_1$ ($K_d = 10 \mu\text{M}$ for SH3_1 binding to $\text{PRM}(\text{H})_1$; Supplementary Fig. 8). Third, we used dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS) to characterize the species that formed during titrations of PRM proteins into SH3_5 (Fig. 2b and Supplementary Figs 10–12). At concentrations below those needed for droplet formation, the addition of either PRM_2 or PRM_4 substantially increased the apparent hydrodynamic radius (R_h , for DLS) and the apparent radius of gyration (R_g , for SAXS), suggesting that oligomeric species formed. For equal total module concentrations, PRM_4 caused larger increases in these values than did PRM_2 , consistent with predictions from polymer theory^{2,3,5,16}. For higher concentrations of PRM proteins, we removed the droplets by centrifugation and found that increasing the concentration of PRM_2 or PRM_4 caused R_h and R_g to decrease in the bulk phase, suggesting that larger species partitioned selectively into the droplet phase. This behaviour resembles that observed in the sol–gel transitions of covalent polymers, in which, above the critical extent of reaction, the average size of oligomers in the

sol phase decreases because larger species preferentially join the gel². Titrations of PRM_1 or $\text{PRM}(\text{H})_1$, which unlike PRM_2 and PRM_4 cannot generate polymers and did not cause phase separation, produced only small, saturable increases in R_h and R_g . Taken together, these data indicate that phase separation is driven by the unique ability of multivalent SH3_m – PRM_n interactions to create large assemblies.

Additional lines of evidence suggest that the multivalent proteins formed large polymers within the droplets, such that the phase transition probably coincides with a sol–gel transition. First, at the extremely high concentrations in the droplet phase, the fractional saturation of SH3 and PRM binding sites is estimated to be threefold to fivefold higher than that needed to induce the sol–gel transition¹⁶ (Supplementary Materials, Cyclization).

Second, DLS analyses of the droplet phases created by mixing SH3_m and PRM_n showed multiphase intensity autocorrelation curves with a complex distribution of relaxation times spanning 0.2–20 ms (Supplementary Fig. 13) or longer (Fig. 2c). Some of these relaxation times nonlinearly scale with the square of the scattering angle (q^2) (Supplementary Fig. 13). The wide range of timescales, the presence of long-timescale processes and the q^2 independence of some of these processes are properties that are typical of polymer solutions in the semi-dilute range but are highly atypical of discrete molecular species¹⁷.

Third, photobleaching experiments showed that the diffusion of the droplet constituents was slowed by about three orders of magnitude relative to free diffusion in water (data not shown). Furthermore, the photobleaching recovery rate correlated inversely with the monomer–monomer affinity and valency (Supplementary Fig. 14), suggesting that recovery represents reorganization of a polymer matrix. Small-molecule fluorophores and enhanced green fluorescent protein (eGFP) can enter, and move rapidly within, the droplets (Supplementary Fig. 15).

Last, cryo-electron microscopy images of SH3_5 plus PRM_5 solutions that were flash-frozen immediately after mixing showed numerous round, electron-dense objects with diameters of about 50–500 nm (Fig. 2d and Supplementary Fig. 16). These objects were not observed with high concentrations of either SH3_5 or PRM_5 alone or when the two components were mixed at concentrations below the droplet concentration. Although the objects were too dense for structures to be observed within them, their edges were highly irregular on the ~10-nm scale, suggesting that they contained large but disordered molecular species, consistent with a crosslinked gel. These characterizations of the droplet phase suggest that this phase contains large polymeric species and has undergone a sol–gel transition.

Taken together, our data suggest a model in which the association of multivalent proteins produces a macroscopic liquid–liquid phase separation, which is thermodynamically coupled to a molecular sol–gel transition within the droplet phase. The sharp transition between small complexes and polymers is consistent with condensation polymerization theory^{2,3,6} and our own particle-based simulations (Supplementary Fig. 5). A theoretical analysis of these systems (Supplementary Information) indicates that the polymerization process is driven appreciably by the extremely high configurational entropy of the polymer^{6,18}.

The behaviour that we observed for these multivalent systems *in vitro* was mirrored in cells. The coexpression of mCherry– SH3_5 and eGFP– PRM_5 fusion proteins in HeLa cells resulted in the formation of approximately 0.5–2-μm diameter cytoplasmic puncta containing both fluorophores (Fig. 3a). Puncta were not observed in cells expressing either protein alone or in cells coexpressing mCherry– SH3_5 and eGFP– PRM_3 , indicating that their formation depends on the interaction between the two high-valency molecules. The puncta did not stain with a large range of vesicle markers or a lipid dye, suggesting that they are phase-separated bodies rather than vesicular structures (Supplementary Fig. 18). Both mCherry and eGFP fluorescence of the bodies recovered within about 10 s of photobleaching (Fig. 3b), indicating that there was rapid exchange of both components with the

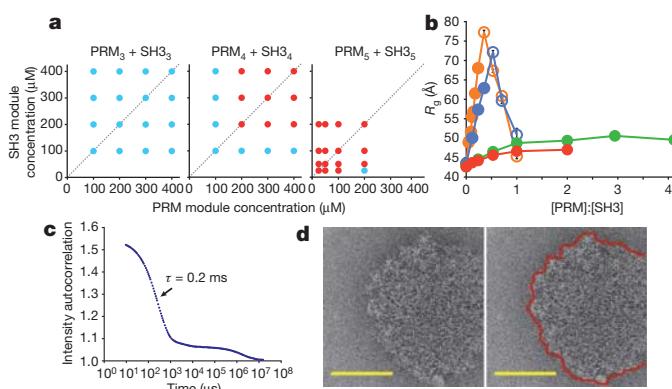


Figure 2 | Multivalency drives phase separation and probably drives a sol–gel transition in the droplet phase. **a**, Phase diagrams of multivalent SH3 and PRM proteins. The concentrations are in terms of the modules. The red circles indicate phase separation, and the blue circles indicate no phase separation. **b**, The R_g values determined from SAXS data that were collected during titrations of PRM proteins into SH3_5 . Closed circles indicate the absence of phase separation; open circles indicate data collected on the supernatant phase, which was separated from the droplets by centrifugation. The titrations used PRM_4 (orange), PRM_2 (blue), PRM_1 (green) and $\text{PRM}(\text{H})_1$ (red). The error bars represent the s.d. calculated from five to ten independent measurements of intensity versus scattering angle (q). **c**, The intensity autocorrelation curve of light scattered at 90° from the pooled droplet phase of SH3_5 plus $\text{PRM}(\text{N}-\text{WASP})_8$. τ , the relaxation time constant of the most rapidly decaying phase. **d**, Cryo-electron microscopy image of a droplet formed by SH3_5 plus PRM_5 (identical image, left and right). The edge of the structure is outlined in red (right). Scale bars, 100 nm.

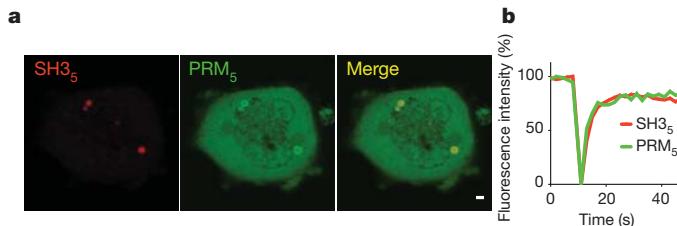


Figure 3 | Coexpression of SH3₅ and PRM₅ in cells produces dynamic puncta. **a**, mCherry-SH3₅ (left), eGFP-PRM₅ (centre) and an image overlay in a cell expressing both proteins (right). It should be noted that the non-uniform eGFP fluorescence in the puncta results from mCherry-eGFP fluorescence resonance energy transfer (FRET) rather than from differential localization of the proteins (Supplementary Fig. 17). Scale bar, 2 μ m. **b**, Both mCherry and eGFP fluorescence recover rapidly after photobleaching.

surrounding cytoplasm and suggesting that the bodies had a dynamic liquid-like nature. Thus, interactions between multivalent proteins can produce phase-separated liquid droplets both *in vitro* and in cells.

The nephrin-NCK-N-WASP system is a natural, three-component interaction that can be used to investigate phase transitions that result from multivalent interactions, as well as the functional consequences of these transitions (Fig. 4a). In kidney podocytes, the transmembrane protein nephrin plays a central role in forming the glomerular filtration barrier, functioning partly through assembling cortical actin¹. The cytoplasmic tail of nephrin contains three tyrosine phosphorylation (pTyr) sites, which can each bind the SH2 domain of NCK^{1,19}. NCK contains three SH3 domains, which can bind the six PRMs in the proline-rich region of N-WASP²⁰. N-WASP, in turn, stimulates the nucleation of actin filaments by the Arp2/3 complex. The multivalency of nephrin or NCK is necessary for proper actin assembly¹⁹ and, together with the multivalency of N-WASP, has the potential to cause phase transitions.

The addition of NCK to an N-WASP construct (GBD-P-VCA, Supplementary Table 1; called N-WASP hereafter) caused droplet formation, as occurred in the model systems described above (Fig. 4b). The addition of a diphosphorylated (2pTyr) nephrin tail peptide dropped the phase boundary for both proteins by more than or equal to twofold (Fig. 4c), presumably because the effective valency of NCK increases when it is arrayed on nephrin. This effect was even more pronounced when nephrin-3pTyr peptide was added (to the same total pTyr concentration) (Fig. 4d). Thus, in cells, protein kinases could regulate phase transitions in this system (and the cooperative assembly of all three proteins) by controlling the degree of phosphorylation of nephrin and, consequently, by shifting the phase boundary from the micromolar (Fig. 4b) to the nanomolar (Fig. 4d) regime. It should be noted that, where measured, the cytoplasmic concentrations of WASP proteins and other actin-regulatory molecules were typically 1–10 μ M, indicating that shifts in this range could be functionally important (Supplementary Fig. 19).

We next asked how droplet formation affects the ability of the nephrin-NCK-N-WASP system to stimulate Arp2/3-mediated actin assembly. We measured the half-time to completion ($t_{1/2}$) of pyrene-actin assembly reactions containing fixed concentrations of the Arp2/3 complex, NCK, nephrin-3pTyr peptide and N-WASP plus variable amounts of an N-WASP truncation mutant (N-WASP Δ) that contains the full proline-rich region and can assemble into polymers but cannot bind the Arp2/3 complex. In the absence of N-WASP Δ , 50 nM N-WASP produced only weak stimulation of the Arp2/3 complex (long $t_{1/2}$) (Fig. 4e). The addition of N-WASP Δ to concentrations between 0 and 750 nM had no effect on actin assembly. However, 1,000 nM N-WASP Δ sharply increased the activation of the Arp2/3 complex. The activity increased asymptotically as the concentration of N-WASP Δ was raised further. N-WASP Δ had no effect on an N-WASP protein that lacks the proline-rich region, suggesting that

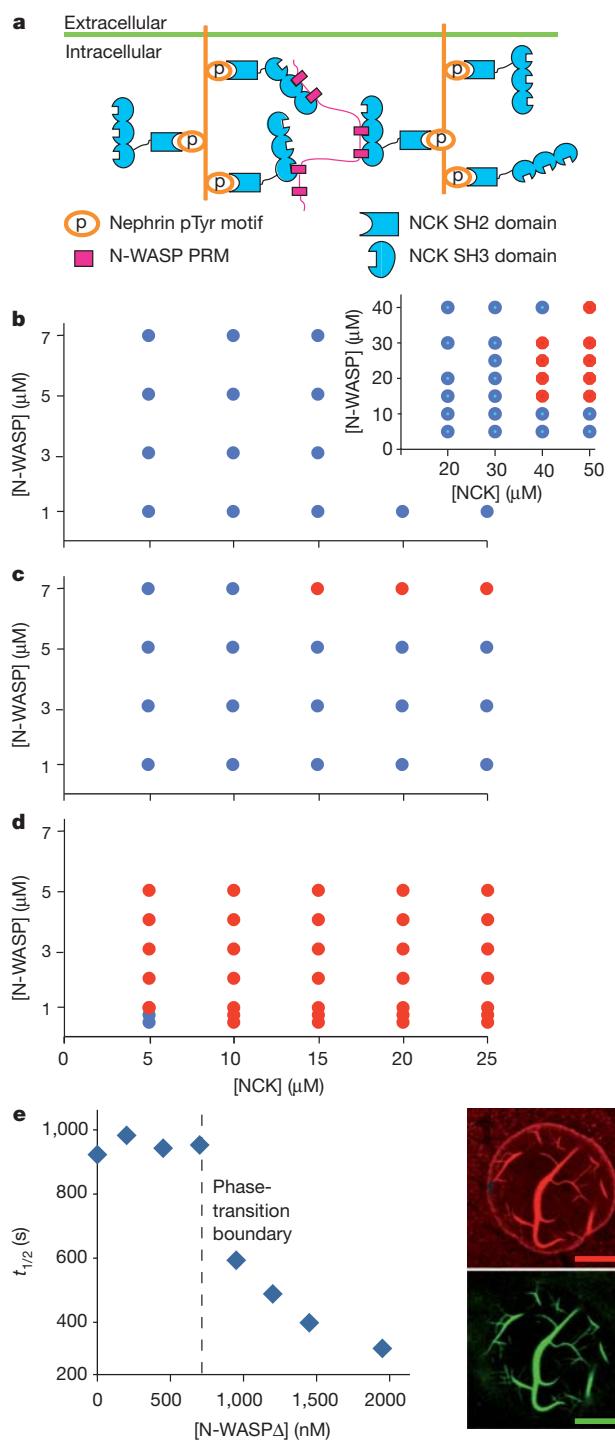


Figure 4 | Phase transition correlates with biochemical activity transition in the nephrin-NCK-N-WASP system. **a**, The interactions of nephrin, NCK and N-WASP. **b–d**, Phase diagrams of N-WASP and NCK alone (**b**, inset shows higher concentrations) or in the presence of 4.5 μ M diphosphorylated nephrin tail peptides (**c**) or 3 μ M triphosphorylated nephrin tail peptides (**d**). The red circles indicate phase separation, and the blue circles indicate no phase separation. **e**, The half-time to completion ($t_{1/2}$) of N-WASP-stimulated actin assembly by the Arp2/3 complex as a function of N-WASP Δ concentration. The vertical dashed line indicates the phase-separation boundary determined in separate assays without actin and the Arp2/3 complex. **f**, Rhodamine-actin (4 μ M, 10% rhodamine-labelled), 300 nM Alexa-488-phalloidin and 10 nM Arp2/3 complex were added to droplets containing triphosphorylated nephrin, NCK and N-WASP, and the droplets were imaged by confocal microscopy; rhodamine (top) and Alexa 488 (bottom) are shown separately. Scale bars, 10 μ m.

engagement by NCK is needed for the increase in activity (Supplementary Fig. 20).

These data are consistent with the switch-like formation of a higher activity^{21,22} (probably polymeric) form of nephrin–NCK–N-WASP when the total concentration of wild-type (active) plus truncated (inactive) N-WASP surpasses that needed for droplet formation. The sharp transition was followed by slower increases in activity as additional N-WASPΔ drew a greater percentage of the wild-type protein into the droplets/polymer. Consistent with these ideas, when mixtures containing N-WASP above its phase-separation concentration were stained with phalloidin, numerous bundles of actin filaments were observed within the droplets (Fig. 4f). In cells, nephrin is a transmembrane protein, and the system would not produce a three-dimensional polymer phase as observed here but rather its two-dimensional equivalent at the plasma membrane. Such interactions might contribute to the formation and/or the stability of the micrometre-scale clusters of nephrin and NCK with the associated actin tails, which are observed in cells when nephrin is crosslinked (and subsequently phosphorylated)¹. Of the 28 known binding partners of the NCK SH2 domain, 14 are predicted or have been demonstrated to contain three or more pTyr sites^{23,24}, suggesting that analogous pTyr–NCK–N-WASP assembly may occur in many systems.

Sharp phase transitions occurring concomitantly with sol–gel transitions may be a general feature of multivalent systems in biology. Many ‘cellular bodies’—that is, subcellular compartments that are compositionally distinct from the surrounding cytoplasm or nucleoplasm but that are not membrane bounded^{25,26}—are enriched in multivalent proteins and nucleic acids^{9,27}. These include promyelocytic leukaemia nuclear bodies, Cajal bodies, P bodies and P granules^{27–29}. Moreover, P granules in the *Caenorhabditis elegans* embryo were recently shown to have liquid-like properties with many of the features that we observed here, including switch-like formation and greatly slowed diffusion of constituent molecules³⁰. Many multivalent proteins, in general, can organize into puncta in the cytosol or at membranes (Supplementary Fig. 21). Within puncta, the physical properties of a polymer could impart micrometre-scale structural and dynamic organization and control the chemistry (for example, of catalysis, molecular interactions or structural rearrangements). Our findings provide a mechanism by which multivalent interactions could yield sharp transitions between physically and functionally distinct states, generating nonlinearity in signalling pathways, connecting disparate length scales in the cell, and perhaps contributing to the structure and function of cellular bodies and other two- and three-dimensional compartments.

METHODS SUMMARY

Material generation. Details are provided in the Supplementary Information.

In vitro phase separation. Samples were incubated for >12 h before scoring them for droplets using bright-field microscopy; when observed, the droplets formed immediately after mixing. For *in vitro* fluorescence recovery after photobleaching (FRAP) experiments, a 5-μm diameter spot was bleached in droplets of >20-μm diameter containing Oregon Green (OG)-labelled SH3₄ (OG–SH3₄) by using a 488-nm laser line. The mean intensity of the bleached spot was fit to a single exponential.

DLS and SAXS. For titrations monitored by DLS (DynaPro, Wyatt) and SAXS (Advanced Photon Source, Biophysics Collaborative Access Team (BioCAT) beamline), the samples contained 170 μM SH3₅ plus PRM proteins at PRM:SH3 module ratios of 0–5. Droplets were removed from all of the relevant samples by centrifugation (at 16,000g for 10 min) before analysis. For SH3₅ plus PRM(N-WASP)₈ and SH3₅ plus PRM₅ droplets, single-angle DLS data were collected on a DynaPro instrument, and multi-angle DLS data were collected on a custom-built apparatus.

Cryo-electron microscopy. The samples were blotted, frozen immediately after mixing and imaged under low-dose cryo conditions in a 2200FS FEG transmission electron microscope (JEOL), using a 2k × 2k slow-scan charge-coupled display (CCD) camera (Tietz).

Cellular assays. HeLa cells were imaged at 32 °C 24 h after transfection with vectors expressing mCherry–SH3₅ and/or eGFP–PRM₅ (or eGFP–PRM₃). Images were acquired on an LSM 510 confocal microscope (Zeiss). For FRAP

experiments, individual puncta containing both mCherry–SH3₅ and eGFP–PRM₅ were bleached with a 488-nm laser line.

Actin polymerization. Actin (4 μM, 5% pyrene-labelled actin) was polymerized in the presence of 3 μM nephrin–3pTyr, 10 μM NCK, 50 nM N-WASP and 10 nM Arp2/3 complex plus increasing concentrations of N-WASPΔ in 150KMEI buffer (150 mM KCl, 1 mM MgCl₂, 1 mM EGTA and 10 mM imidazole, pH 7) as described previously²¹. Imaging was performed on an LSM 510 confocal microscope after adding rhodamine–actin (4 μM, 10% rhodamine-labelled actin), 300 nM Alexa 488–phalloidin and 10 nM Arp2/3 complex to droplets containing 3 μM nephrin–3pTyr, 10 μM NCK and 2 μM N-WASP.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 May 2010; accepted 20 January 2012.

Published online 7 March 2012.

1. Jones, N. *et al.* Nck adaptor proteins link nephrin to the actin cytoskeleton of kidney podocytes. *Nature* **440**, 818–823 (2006).
2. Flory, P. J. *Principles of Polymer Chemistry* (Cornell Univ. Press, 1953).
3. Cohen, R. J. & Benedek, G. B. Equilibrium and kinetic theory of polymerization and the sol–gel transition. *J. Phys. Chem.* **86**, 3696–3714 (1982).
4. Lehn, J.-M. Supramolecular polymer chemistry—scope and perspectives. *Polym. Int.* **51**, 825–839 (2002).
5. Tanaka, F. *Polymer Physics: Applications to Molecular Association and Thermoreversible Gelation* (Cambridge Univ. Press, 2011).
6. Semenov, A. N. & Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules* **31**, 1373–1385 (1998).
7. Brewer, C. F., Miceli, M. C. & Baum, L. G. Clusters, bundles, arrays and lattices: novel mechanisms for lectin–saccharide-mediated cellular interactions. *Curr. Opin. Struct. Biol.* **12**, 616–623 (2002).
8. Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452 (2003).
9. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nature Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
10. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nature Rev. Mol. Cell Biol.* **8**, 983–994 (2007).
11. Goldberg, R. A theory of antibody–antigen reactions. I. Theory for reactions of multivalent antigen with bivalent and univalent antibody. *J. Am. Chem. Soc.* **74**, 5715–5725 (1952).
12. Dam, T. K. *et al.* Thermodynamic, kinetic, and electron microscopy studies of concanavalin A and *Dioclea grandiflora* lectin cross-linked with synthetic divalent carbohydrates. *J. Biol. Chem.* **280**, 8640–8646 (2005).
13. Sisu, C. *et al.* The influence of ligand valency on aggregation mechanisms for inhibiting bacterial toxins. *ChemBioChem* **10**, 329–337 (2009).
14. Jin, J. *et al.* Eukaryotic protein domains as functional units of cellular evolution. *Sci. Signal.* **2**, ra76 (2009).
15. Asherie, N. *et al.* Oligomerization and phase separation in globular protein solutions. *Biophys. Chem.* **75**, 213–227 (1998).
16. Stockmayer, W. H. Molecular distribution in condensation polymers. *J. Polym. Sci.* **9**, 69–71 (1952).
17. Li, J., Ngai, T. & Wu, C. The slow relaxation mode: from solutions to gel networks. *Polym. J.* **42**, 609–625 (2010).
18. Semenov, A., Charlot, A., Auzely-Velty, R. & Rinaudo, M. Rheological properties of binary associating polymers. *Rheol. Acta* **46**, 541–568 (2007).
19. Blasutig, I. M. *et al.* Phosphorylated YDXV motifs and Nck SH2/SH3 adaptors act cooperatively to induce actin reorganization. *Mol. Cell. Biol.* **28**, 2035–2046 (2008).
20. Rohatgi, R., Nollau, P., Ho, H. Y., Kirschner, M. W. & Mayer, B. J. Nck and phosphatidylinositol 4,5-bisphosphate synergistically activate actin polymerization through the N-WASP–Arp2/3 pathway. *J. Biol. Chem.* **276**, 26448–26452 (2001).
21. Padrick, S. B. *et al.* Hierarchical regulation of WASP/WAVE proteins. *Mol. Cell* **32**, 426–438 (2008).
22. Padrick, S. B. & Rosen, M. K. Physical mechanisms of signal integration by WASP family proteins. *Annu. Rev. Biochem.* **79**, 707–735 (2010).
23. Lettau, M., Pieper, J. & Janssen, O. Nck adaptor proteins: functional versatility in T cells. *Cell Commun. Signal.* **7**, 1 (2009).
24. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. ScanSite 2.0: proteome-wide prediction of cell signalling interactions using short sequence motifs. *Nucleic Acids Res.* **31**, 3635–3641 (2003).
25. Matera, A. G., Izaguirre-Sierra, M., Praveen, K. & Rajendra, T. K. Nuclear bodies: random aggregates of sticky proteins or crucibles of macromolecular assembly? *Dev. Cell* **17**, 639–647 (2009).
26. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36**, 932–941 (2009).
27. Parker, R. & Sheth, U. P. Bodies and the control of mRNA translation and degradation. *Mol. Cell* **25**, 635–646 (2007).
28. Matera, A. G. & Shpargel, K. B. Pumping RNA: nuclear bodybuilding along the RNP pipeline. *Curr. Opin. Cell Biol.* **18**, 317–324 (2006).
29. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nature Rev. Mol. Cell Biol.* **8**, 1006–1016 (2007).

30. Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Onuchic and S. Padrick for discussion of the theoretical aspects of this study, L. Rice for sharing his fluorescence microscope, M. Socolich for a gift of purified eGFP, K. Luby-Phelps and A. Bugde for advice on FRAP experiments, S. Padrick and L. Doolittle for help in purifying actin and the Arp2/3 complex and for sharing reagents, N. Grishin and S. Shi for help with database searches, K. Lynch for providing the PTB expression construct, D. Billadeau and T. Gomez for providing antibodies, A. Ramesh, W. Winkler and P.-L. Tsai for advice on RNA experiments, K. Roybal and C. Wülfing for sharing unpublished data, and J. Liu for help with cryo-electron tomography. This work was supported by the following: the Howard Hughes Medical Institute and grants from the National Institutes of Health (NIH) (R01-GM56322) and Welch Foundation (I-1544) to M.K.R., a Chilton Foundation Fellowship to H.-C.C., an NIH EUREKA award (R01-GM088745) to Q.-X.J., an NIH Cancer Biology T32 Training Grant to M.L., a National Science Foundation award (DMR-1005707) to P.S.R. and a Gates Millennium Fund award to J.V.H. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic

Energy Sciences, Office of Science, under contract number W-31-109-ENG-38. BioCAT is NIH-supported Research Center RR-08630.

Author Contributions M.K.R. oversaw the project, helped analyse all of the data and wrote the paper with assistance from all of the authors. P.L., H.-C.C. and M.K.R. conceived of the project. P.L. developed and interpreted the theoretical and computational models, which promoted much of the experimentation. S.B. performed and analysed experiments on the nephrin–NCK–N–WASP system and performed monovalent competition studies. H.-C.C. mapped and analysed the phase diagrams, and collected FRAP data, on the engineered model systems. S.K. performed and analysed the cellular experiments. S.B., B.C., L.G. and B.T.N. collected and/or analysed the SAXS data. S.B., M.L. and Q.-X.J. collected and/or analysed the electron microscopy data. S.B., J.V.H. and P.S.R. collected and/or analysed the multi-angle DLS data. H.-C.C. and S.B. collected and analysed the single-angle DLS data. D.S.K. synthesized the octameric PRM dendrimer. S.F.B. analysed the cyclization in the sol–gel transition.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K.R. (Michael.Rosen@utsouthwestern.edu).

METHODS

Measurements of protein stoichiometry and concentration in droplets. All phase-separation assays were performed in 150KMEI buffer (150 mM KCl, 1 mM MgCl₂, 1 mM EGTA and 10 mM imidazole, pH 7), except where indicated. SH3₄ and PRM₄ (doped with 0.5% of either protein labelled with Oregon Green) were incubated at a 1:1 stoichiometry at various concentrations at room temperature overnight. The droplet phase (where formed) was collected by centrifugation, and the supernatant was removed and digested with proteinase K (Promega) overnight. The fluorescence intensity (excitation wavelength (λ_{ex}) = 488 nm, and emission wavelength (λ_{em}) = 512–521 nm) of the samples was converted to concentration using a standard curve of Oregon Green alone. The data revealed a 1:1 stoichiometry in all cases in both the droplet and the supernatant phases. In other systems that were mixed at a 1:1 module ratio, we calculated the concentrations from the absorbance at 280 nm using module-weighted extinction coefficients, assuming a 1:1 module stoichiometry in all phases.

SAXS. SAXS experiments were performed at the Biophysics Collaborative Access Team (BioCAT) undulator beamline 18-ID at the Advanced Photon Source, Argonne National Laboratory as described previously³¹. All samples were centrifuged for 10 min at 16,000g at 4 °C before mixing. In mixtures above the phase-transition concentration, solutions were centrifuged for an additional 10 min to remove the droplet phase before SAXS analysis. To minimize radiation damage, all samples were in 150KMEI buffer containing 5 mM TCEP and were flowed at 5 $\mu\text{l s}^{-1}$ during data collection³¹. Scattering intensity profiles over the q range from 0.006 to 0.36 \AA^{-1} were calculated from radial averaging of the two-dimensional scattering patterns using macros written by the BioCAT staff for the software IGOR Pro (WaveMetrics). The scattering for each sample and the accompanying buffer were averaged from five to ten repeats, and the protein scattering was obtained by subtracting the buffer scattering. Guinier plots were used to calculate R_g and to ascertain the absence of aggregates. Alternatively, R_g and the distance distribution function, $P(r)$, were calculated using the program GNOM³². The apparent molecular weight was estimated³³ by integrating the Kratky plot to a maximum q at 0.2 \AA^{-1} .

DLS titration. In the titration experiments, scattering measurements were performed on a DynaPro instrument (Wyatt) at 30% laser power and 22 °C. Twenty repetitive data sets, each of 100 s duration, were averaged for each titration point. Autocorrelation data were analysed using the regularization method in Dynamics version 6.4.3 software (Proterion Corporation). All samples were centrifuged before and after mixing, as for the SAXS titration.

DLS of SH3₅ plus PRM(N-WASP)₈ droplets. To collect enough volume of the droplet phase (~25 μl), 850 μM SH3₅ was mixed with 850 μM PRM(N-WASP)₈ (module concentrations) to a final volume of 900 μl in 150KMEI buffer. The denser, droplet, phase was collected by centrifugation (16,000g for 10 min) and analysed using a DynaPro instrument at 22 °C and 20% laser power. Twenty repetitive data sets, each of 1,000 s duration, were averaged.

Multi-angle DLS of SH3₅ plus PRM₅ droplet and supernatant phases. SH3₅ (850 μM) was mixed with PRM₅ (850 μM) (module concentrations) to a final volume of 900 μl to collect enough volume of the droplet phase (~20 μl). The denser, droplet, phase was collected by centrifugation. Measurements were made using a custom-built apparatus equipped with an Innova 90 argon laser (Coherent) set to 514.5 nm. A wide-range photometer-preamplifier-discriminator (Pacific Precision Instruments) drove a pulse shaper (ALV), which fed an ALV-5000 digital autocorrelator (ALV). The temperature was maintained at 25 °C by a circulating

water bath. For the supernatant phase, three 10-min runs were collected and averaged at each scattering angle between 30° and 110°. For the droplet phase, a single 30-min run was collected at each angle. The distributions of the decay rate and the average decay rate were determined by analysing the correlation functions using an inverse Laplace transformation in the program CONTIN³⁴.

Cryo-electron microscopy. Carbon-coated copper grids were glow-discharged in an EMS-100 unit with a 40 mA current for 90 s. SH3₅ (100 μM) was loaded onto the grid, followed by PRM₅ (100 μM) (module concentrations), giving a 4 μl final volume. The mixture was immediately blotted for 6.5 s in a Vitrobot Mark III (Gatan) before being plunged into liquid ethane. Mixtures of SH3₅ and PRM₅ below the critical concentrations (2.5 μM module concentration each), or SH3₅ or PRM₅ alone (250 μM module concentration) were prepared identically. Samples contained 150KMEI buffer. Frozen grids were stored in liquid nitrogen until electron microscopy analysis. Imaging was performed under low-dose cryo conditions in a 2200FS FEG transmission electron microscope (JEOL), using a 2k \times 2k slow-scan charge-coupled display (CCD) camera (Tietz). The dose rate was adjusted to ~20 electrons $\text{\AA}^{-2} \text{s}^{-1}$; the defocus level varied from -1.0 to -3.0 μm .

Cell culture and fluorescence microscopy. HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% FBS. Cells were transfected using Effectene (QIAGEN) and incubated for 24 h at 37 °C. For Nile-Red staining, cells were incubated with Nile Red (at a final concentration of 0.5 $\mu\text{g ml}^{-1}$) for 15, 30 or 60 min before imaging. For immunofluorescence analysis, transfected cells were fixed with 4% paraformaldehyde in PBS and permeabilized with 0.1% Triton X-100 in PBS for 5 min. The antibodies used for staining were anti-EEA1, anti-GM130, anti-golgin-97, anti-VPS35, anti-CIMPR, anti-clathrin heavy chain, anti-LAMP1, anti-SNX1, anti-SNX2, anti-TGN46 (provided by D. Billadeau and T. Gomez) and anti-caveolin-1 antibodies. Secondary antibodies conjugated to Alexa Fluor 647 were obtained from Invitrogen. Images were collected on a DeltaVision system (Applied Precision). The mCherry-SH3₅ and eGFP-PRM₅ constructs were generated by cloning SH3₅ and PRM₅ into the vectors pmCherry-N1 and peGFP-N1, respectively. Images were processed using the program ImageJ (<http://rsb.info.nih.gov/ij>).

In vitro FRAP assays. Fluorescence recovery after photobleaching (FRAP) experiments were performed on an LSM 510 confocal microscope (Zeiss) using a 488-nm laser line. The OG-SH3₄ or eGFP was bleached using ten iterative pulses (total time ~0.55 s) with full laser power. The droplets were >20 μm in diameter, and the bleached spots were 5 μm in diameter. The imaging power was 0.5% for OG-SH3₄ and 2% for eGFP. Images were processed using ImageJ. Background-corrected data were fit to a single exponential decay to yield the recovery time constant using GraphPad Prism 5 (GraphPad Software).

- Chen, B. *et al.* ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, σ 54. *Structure* **15**, 429–440 (2007).
- Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
- Fischer, H., Neto, M. O., Napolitano, H. B., Polikarpov, I. & Craievich, A. F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.* **43**, 101–109 (2010).
- Provencher, S. W. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* **27**, 229–242 (1982).

A petunia ABC protein controls strigolactone-dependent symbiotic signalling and branching

Tobias Kretzschmar¹, Wouter Kohlen^{2*}, Joelle Sasse^{1*}, Lorenzo Borghi¹, Markus Schlegel¹, Julien B. Bachelier¹, Didier Reinhardt³, Ralph Bours², Harro J. Bouwmeester^{2,4} & Enrico Martinoia¹

Strigolactones were originally identified as stimulators of the germination of root-parasitic weeds¹ that pose a serious threat to resource-limited agriculture². They are mostly exuded from roots and function as signalling compounds in the initiation of arbuscular mycorrhizae³, which are plant-fungus symbionts with a global effect on carbon and phosphate cycling⁴. Recently, strigolactones were established to be phytohormones that regulate plant shoot architecture by inhibiting the outgrowth of axillary buds^{5,6}. Despite their importance, it is not known how strigolactones are transported. ATP-binding cassette (ABC) transporters, however, are known to have functions in phytohormone translocation^{7–9}. Here we show that the *Petunia hybrida* ABC transporter PDR1 has a key role in regulating the development of arbuscular mycorrhizae and axillary branches, by functioning as a cellular strigolactone exporter. *P. hybrida* *pdr1* mutants are defective in strigolactone exudation from their roots, resulting in reduced symbiotic interactions. Above ground, *pdr1* mutants have an enhanced branching phenotype, which is indicative of impaired strigolactone allocation. Overexpression of *Petunia axillaris* PDR1 in *Arabidopsis thaliana* results in increased tolerance to high concentrations of a synthetic strigolactone, consistent with increased export of strigolactones from the roots. PDR1 is the first known component in strigolactone transport, providing new opportunities for investigating and manipulating strigolactone-dependent processes.

Strigolactones are a new class of carotenoid-derived¹⁰ phytohormone in land plants. In addition to their role in shoot branching, strigolactones are exuded into the rhizosphere under phosphorus-limiting conditions⁵ and act as growth stimulants of arbuscular mycorrhizal fungi³. To identify efflux carriers of arbuscular-mycorrhiza-promoting factors such as strigolactones, we used a degenerate primer approach (Supplementary Fig. 2a) to isolate full-size PDR-type transporters (also known as ABC subtype G (ABCG) transporters) of *P. hybrida* that are abundant in phosphate-starved or mycorrhizal roots. The rationale behind the focus on these transporters, of which there are 15 in *Arabidopsis*¹¹, 23 in *Oryza sativa* (rice)¹¹ and 23 putative factors in *Solanum lycopersicum* (tomato) (Supplementary Fig. 3a), was that they are plasma membrane proteins often found in roots¹², they are implicated in below-ground plant-microbe interactions^{13,14}, and they have affinities for compounds that are structurally related to strigolactones^{8,9,15}. Of six primary candidates, only *P. hybrida* PDR1 had increased expression in roots that were subjected to either phosphate starvation (Fig. 1a) or colonization by the arbuscular mycorrhizal fungus *Glomus intraradices* (Fig. 1b). Furthermore, PDR1 transcript levels increased in response to treatment with the synthetic strigolactone analogue GR24 or the auxin analogue 1-naphthaleneacetic acid (NAA) (Fig. 1c). Auxin has been shown to upregulate strigolactone-biosynthesis genes¹⁶ and to be involved in pre-symbiotic and early mycorrhizal events¹⁷.

P. hybrida PDR1 is predicted to encode a full-size PDR cluster I protein (GenBank accession number JQ292813; Supplementary Figs

2b–c and 3b). The closest *Arabidopsis* homologue, ABCG40 (also known as PDR12), transports abscisic acid (ABA)⁹. However, in contrast to *Arabidopsis* ABCG40, *P. hybrida* PDR1 is not regulated by ABA (Fig. 1c). A 1.8-kilobase (kb) element upstream of PDR1 (GenBank

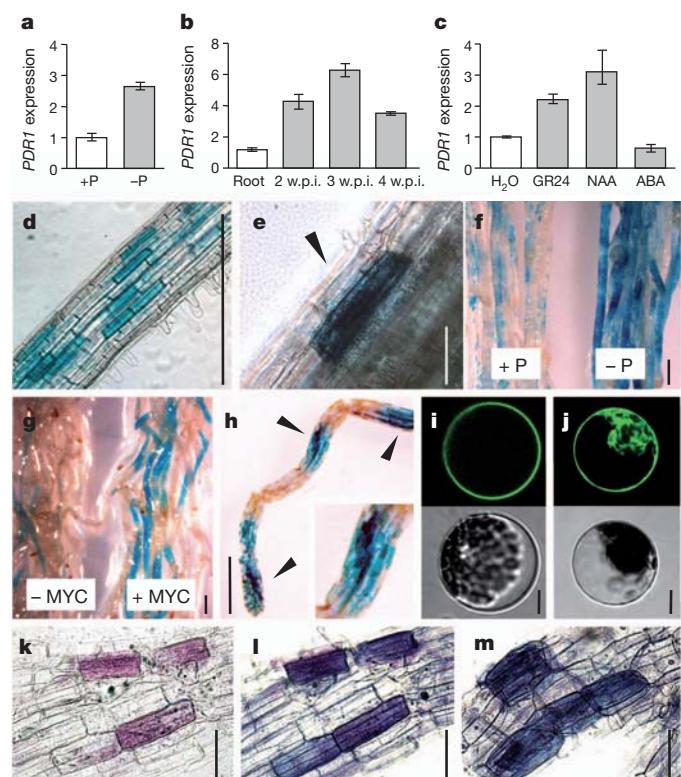


Figure 1 | Below-ground PDR1 expression and PDR1 localization. a–c, Quantitative PCR for PDR1 in W115 roots: in the presence or absence of phosphate (P) (a); at 2–4 weeks post inoculation (w.p.i.) with *G. intraradices* (b); and in response to water (H₂O), GR24, NAA or ABA (c). Data are presented as mean \pm s.e.m. ($n = 3$, where n denotes the number of samples). d–h, The *pPhPDR1::GUS* signal in W115 roots without treatment (d, e) (black arrowheads indicate epidermal cells); under phosphate-sufficient and phosphate-deficient conditions (f); in response to mycorrhization (+MYC, 8 w.p.i.) (g); and in mycorrhized roots (8 w.p.i.) costained with black ink (h). Scale bars, 1 mm (d, f–h) and 0.1 mm (e). i, j, Transient CaMV 35S::GFP::gPaPDR1 expression in *Arabidopsis* mesophyll protoplasts. The GFP::gPaPDR1 signal (i, top) and the corresponding transmission electron micrograph (i, bottom) and the free GFP signal (j, top) and transmission image (j, bottom) are shown. Scale bars, 10 μ m. k–m, *pPhPDR1::GUS* signal co-localization with trypan-blue-stained root HPCs. Shown are a magenta GUS-stained root section (k), an additional trypan blue stain of the same sample (l) and stained wild-type root HPCs (m). Scale bars, 0.1 mm.

¹Institute of Plant Biology, University of Zurich, 8008 Zurich, Switzerland. ²Laboratory of Plant Physiology, Wageningen University, 6700 AR Wageningen, The Netherlands. ³Department of Biology, University of Fribourg, 1700 Fribourg, Switzerland. ⁴Centre for Biosystems Genomics, PO Box 98, 6700 AB Wageningen, The Netherlands.

*These authors contributed equally to this work.

accession number JQ292814) was fused to the GUS reporter and stably transformed into the *P. hybrida* cultivar W115. Below ground, *pPhPDR1:GUS* (promoter *P. hybrida* *PDR1*–GUS construct) expression was substantial in individual subepidermal cells of the lateral roots (Fig. 1d, e). These cells were largely identical to hypodermal passage cells (HPCs) (Fig. 1k–m), which are devoid of suberin and serve as cortical entry points for arbuscular mycorrhizal hyphae¹⁸. GUS staining was more prominent in roots grown under phosphate-deficient conditions (Fig. 1f) and in mycorrhizal roots, particularly in regions containing or flanking fully developed arbuscular mycorrhizal structures (Fig. 1g, h). These results suggested a role for PDR1 in arbuscular mycorrhiza during pre-symbiotic development and during intraradical colonization. To assess the localization of the PDR1 protein, the genomic *PDR1* orthologue from the *P. hybrida* progenitor *P. axillaris* (which has 99.7% amino acid identity to *P. hybrida* PDR1; GenBank accession number JQ292812; Supplementary Fig. 2b) was fused to the carboxy-terminus-encoding region of the green fluorescent protein (GFP) gene (*GFP:gPaPDR1*). Transient expression of *GFP:gPaPDR1* in *Arabidopsis* showed that PDR1 localizes to the plasma membrane (Fig. 1i, j), consistent with a role in secretion.

For functional analysis, we screened the dTph1 transposon-containing *P. hybrida* line W138 for insertional *pdr1* mutants. A PCR-based DNA library screen of 1,000 individuals led to the identification of a dTph1 insertion in exon 4 of *P. hybrida* *PDR1* (Supplementary Fig. 4a–d), together with a footprint allele causing a frame shift (Supplementary Fig. 4e). Insertion of dTph1 in the coding region of a gene frequently results in a complete loss of function¹⁹.

The W138 line with the mutant *pdr1* gene (W138-*pdr1*) was compared directly with W138 and was crossed with *P. hybrida* W115 for further segregation analysis. Five homozygous *pdr1* mutant lines (W115 × W138-*pdr1*) and wild-type lines (W115 × W138) were derived from the F₂ generation (Supplementary Fig. 4d). Phenotypes cosegregated with the mutation in *PDR1*, and transposon display analysis did not uncover other cosegregating dTph1 insertions in the W115 × W138 lines (Supplementary Fig. 4f), suggesting that dTph1 insertion in *PDR1* is responsible for the observed phenotypes. In addition, *P. hybrida* *PDR1* knockdown lines (*pdr1*-RNAi), which were created in W115 by using two independent RNA interference (RNAi) constructs (Supplementary Fig. 5a, b), showed similar phenotypes.

W138-*pdr1* had a significantly reduced ability to accommodate *Gigaspora margarita* and *G. intraradices* (Fig. 2a), two distantly related arbuscular mycorrhizal fungi with different growth strategies⁴. This finding indicated that PDR1 functions as a transporter of a stimulatory molecule involved in symbiosis with diverse arbuscular mycorrhizal fungal species. Indeed, root exudates from W138-*pdr1* showed reduced activity for stimulating the hyphal branching of *G. margarita* in an *in vitro* bioassay (Fig. 2b). Because these results suggested an involvement of strigolactones, W115 × W138-*pdr1* root exudates were assessed for their ability to stimulate the germination of the root-parasitic weed *Phelipanche ramosa* (Orobanchaceae). As a control, root exudates of *dad1* mutants from the *P. hybrida* cultivar V26 were used. *DAD1* encodes carotenoid cleavage dioxygenase 8 (CCD8)²⁰ and is an orthologue of the established strigolactone-biosynthesis genes *RMS1*, *MAX4* and *D10* in *Pisum sativum* (pea), *Arabidopsis* and rice, respectively^{5,6}. The germination rate of *P. ramosa* was significantly lower in the presence of root exudates from W115 × W138-*pdr1* or *dad1* mutants than root exudates from the corresponding wild-type plants, which induced germination to a similar extent to GR24 (Fig. 2d). Comparable results were obtained with *pdr1*-RNAi lines (Supplementary Fig. 5c). When inoculated with *G. intraradices*, W115 × W138-*pdr1* lines had similarly retarded colonization rates to W138-*pdr1* and *dad1* mutants (Fig. 2c).

Despite the delay in arbuscular mycorrhizal development, neither W115 × W138-*pdr1* nor *dad1* mutants had any morphological aberrations in intracellular mycorrhizal structures (Fig. 2e–h). Intraradical hyphae and arbuscules appeared normal, suggesting that the quantitative

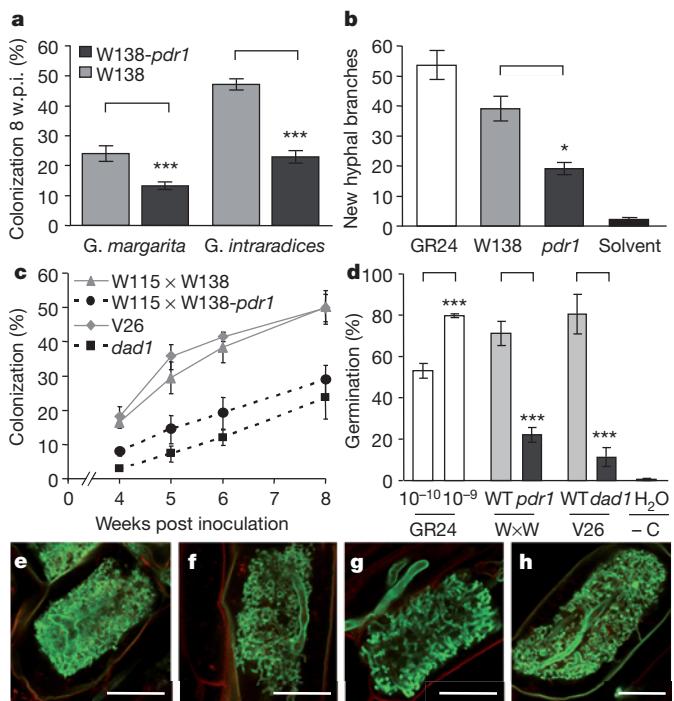


Figure 2 | Below-ground *pdr1* phenotypes. **a**, Arbuscular mycorrhizal colonization of W138 and W138-*pdr1* roots 8 w.p.i. with two arbuscular mycorrhizal fungi. Data are presented as mean \pm s.e.m. ($n > 20$). **b**, *In vitro* branching response of *G. margarita* 24 h after application of GR24 ($n = 5$), root exudates from W138 ($n = 17$), root exudates from W138-*pdr1* ($n = 36$) or 10% acetone (solvent, $n = 5$). Data are presented as mean \pm s.e.m. **c**, Kinetics of *G. intraradices* colonization of W115 × W138 ($n = 25$), W115 × W138-*pdr1* ($n = 25$), V26 ($n = 5$) and *dad1* mutants ($n = 5$). Data are presented as mean \pm s.e.m. ($P < 0.001$ for all time points between mutants and wild types). **d**, *P. ramosa* germination induced by GR24 ($n = 3$), root exudates from W115 × W138 (W×W WT, $n = 10$), W115 × W138-*pdr1* (W×W *pdr1*, $n = 10$), V26 ($n = 4$), *dad1* mutants ($n = 4$) or water (H_2O , $n = 4$). Data are presented as mean \pm s.e.m. – C, negative control. **e–h**, *G. intraradices* intracellular arbuscular mycorrhizal morphology 4 w.p.i. in W115 × W138 (e), W115 × W138-*pdr1* (f), V26 (g) and *dad1* mutants (h). Scale bars, 20 μ m. **a–d**, * , $P < 0.05$; *** , $P < 0.001$.

differences in colonization were due to a decreased number of hyphal penetrations and retarded intraradical expansion of arbuscular mycorrhizal fungal colonies, rather than to defects in intracellular fungal development. Thus, the phenotype of *dad1* and *pdr1* mutants was distinct from that of arbuscular mycorrhizal mutants such as *pam1* (ref. 21), *str1* (ref. 22) or SYM-pathway⁴ mutants, which commonly have aberrant arbuscular mycorrhizal fungal structures.

A detailed analysis of W138 root exudates resulted in the identification of the strigolactone orobanchol (Supplementary Fig. 6). The orobanchol levels in *pdr1* plant root exudates were significantly lower than those in wild-type plant exudates (Fig. 3a), whereas the levels in root extracts were not affected (Fig. 3b), indicating that *pdr1* mutants are not defective in strigolactone biosynthesis. Orobanchol was not detectable in root exudates or in root extracts of *dad1* plants, confirming this mutant's purported defect in strigolactone biosynthesis (Fig. 3a, b). The finding that only extraradical orobanchol levels were affected in *pdr1* mutants indicated that PDR1 functions as a strigolactone export carrier.

PDR1-dependent strigolactone transport was further explored in a heterologous system by stable and constitutive overexpression of *GFP:gPaPDR1* in *Arabidopsis* Col-0, resulting in *PDR1*-OE lines (Supplementary Fig. 7a, b). *Arabidopsis* does not form arbuscular mycorrhizae and exudes only minuscule quantities of strigolactones²³. When grown on GR24-containing medium, *PDR1*-OE lines proved

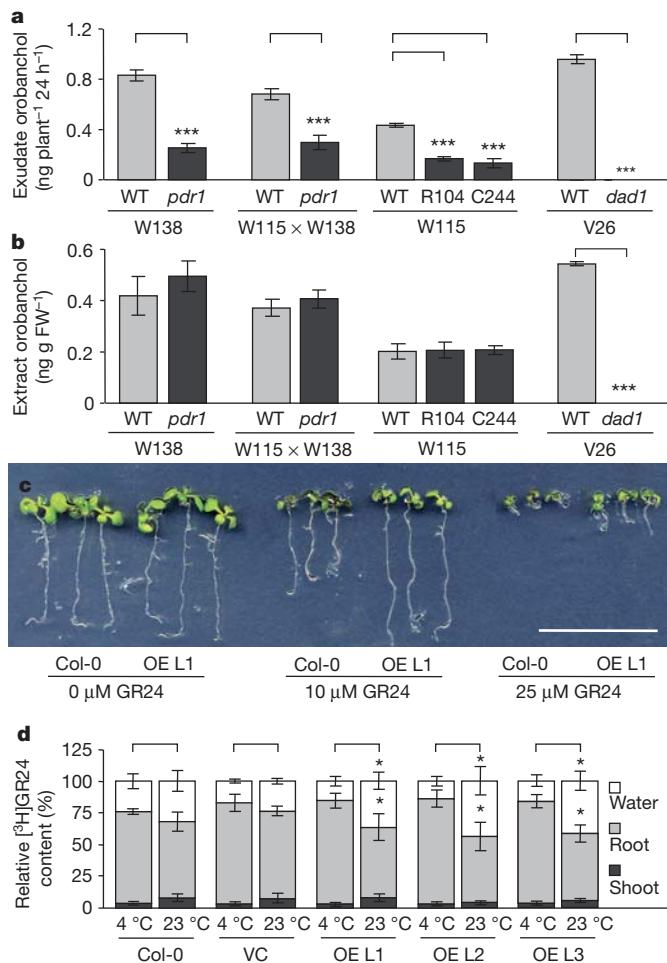


Figure 3 | Orobanthol content and PDR1-dependent GR24 tolerance and transport. **a, b**, Orobanthol in the root exudates (a) and extracts (b) of *pdr1* lines, *dad1* mutants and wild-type plants ($n = 9$). Data are presented as mean \pm s.e.m. FW, fresh weight. **c**, Col-0 and *PDR1*-OE lines grown on 0, 10 and 25 μ M GR24. Scale bar, 1 cm. **d**, GR24 export assay of [3 H]GR24-preloaded roots of Col-0, vector control (VC) and *PDR1*-OE lines (OE L1–3). Relative [3 H]GR24 in the medium (water), root and shoot after 1 h incubation at 4 °C and 23 °C. Data are presented as mean \pm s.e.m. ($n = 8$). **a, b, d**, $^*, P < 0.05$; $^{***}, P < 0.001$.

more tolerant to the deleterious effects of high strigolactone concentrations on root elongation²⁴ than did wild-type plants (Fig. 3c and Supplementary Fig. 7c). Direct strigolactone exudation was assessed by quantifying the efflux of preloaded [3 H]GR24 from roots either incubated at 4 °C, to monitor passive diffusion, or at 23 °C, allowing transporter-mediated efflux. After 1 h, *PDR1*-OE roots incubated at 23 °C retained significantly less GR24 than *PDR1*-OE roots incubated at 4 °C (Fig. 3d and Supplementary Fig. 7d). In agreement with this observation, more GR24 was found in the root exudates of *PDR1*-OE lines at 23 °C. No significant differences were found between the root extracts and root exudates of wild-type or vector control lines in either condition (Fig. 3d). These results, together with the observed GR24-resistance phenotype of *PDR1*-OE plants, are best explained by *PDR1* acting as a strigolactone exporter.

Taken together, our data suggest a role for *PDR1* in strigolactone secretion from HPCs. We propose that *PDR1*-mediated strigolactone exudation under low phosphate conditions creates local rhizosphere gradients that guide arbuscular mycorrhizal hyphae to HPCs, which are susceptible to hyphal penetration (Supplementary Fig. 1), thereby initiating arbuscular mycorrhizae. The symbiotic phenotype of *pdr1* and *dad1* mutants, and the induction of *PDR1* expression in colonized root segments, suggests that strigolactones play an additional role in

promoting sustained intercellular root colonization, whereas intracellular stages (for example, arbuscules) develop independently of strigolactones (Fig. 2e–h).

Recently, it was demonstrated that strigolactones inhibit shoot branching^{5,6}. Although certain aspects of strigolactone biosynthesis and signalling have been unravelled²⁵, information about the mode of strigolactone transport is scant. Strigolactones are mobile within the xylem sap²³, but it is not known how they are released from the producing cells and whether directed cell-to-cell transport occurs. Strigolactone biosynthesis is subject to direct negative feedback regulation²⁶. Hence, strigolactone biosynthesis needs to be coordinated with export to prevent strigolactones from accumulating to levels that would restrict further production. Indeed, *PDR1* expression was found to be stimulated by exogenous application of GR24 (Figs 1c and 4a), suggesting that this gene is induced in a substrate-dependent manner, as previously observed for other PDR subfamily^{9,15} members.

Above ground, *PDR1* expression was largely confined to stem tissues, particularly the vasculature and nodal tissues adjacent to leaf axils (Fig. 4b, c). However, *PDR1* expression was absent from dormant buds (Fig. 4d). This pattern is consistent with *PDR1* functioning as

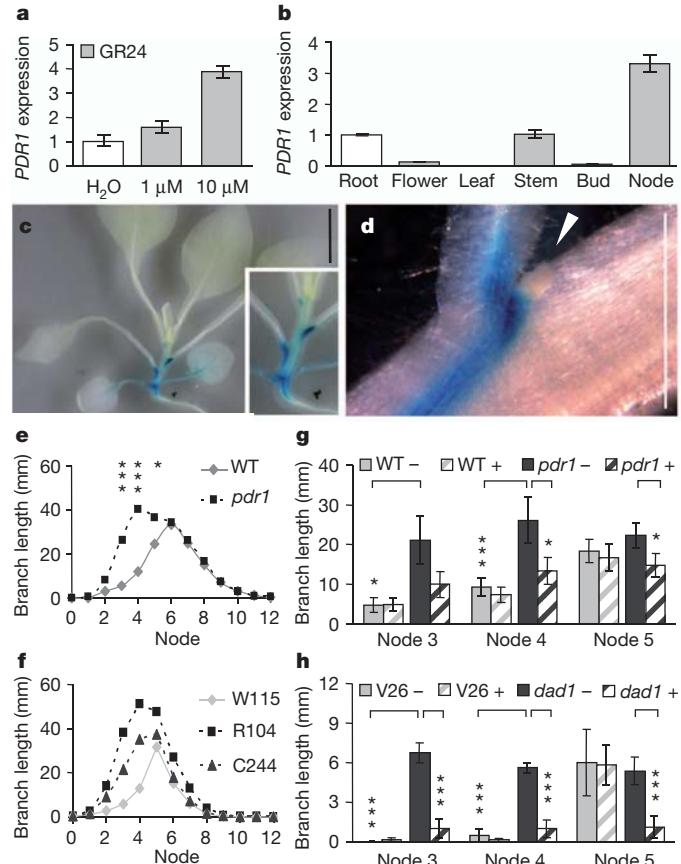


Figure 4 | Above-ground *PDR1* expression and *pdr1*-related branching phenotypes. **a, b**, Quantitative PCR for above-ground W115 GR24-treated tissue (a) and various organs (b). Data are presented as mean \pm s.e.m. ($n = 3$). **c, d**, *pPhPDR1:GUS* in W115 at the four-leaf stage (c) and in the node (d) (white arrowhead, dormant axillary bud). The inset in c shows a magnification of the lower right corner of c. Scale bars, 10 mm (c) and 1 mm (d). **e, f**, Branch development 41 days post germination. Data are presented as mean \pm s.e.m. Branch length is shown for W115 \times W138 and W115 \times W138-*pdr1* ($n = 110$) (e) and for W115 and two *pdr1*-RNAi lines, R104 ($P < 0.001$ for node 3–4) and C244 ($P < 0.05$ for node 3) ($n = 8$) (f). **g, h**, Effects of GR24 (+, striped bars) on branch development 34 days post germination at node 3–5 for W115 \times W138 (WT) and W115 \times W138-*pdr1* (*pdr1*) ($n = 24$) (g) and for V26 and *dad1* mutants ($n = 8$) (h). Data are presented as mean \pm s.e.m. **a, b, e, h**, $^*, P < 0.05$; $^{***}, P < 0.001$.

a strigolactone transporter. Although strigolactones are mobile in the xylem²³, a cellular transport system is required to deliver strigolactones to dormant buds that are not yet connected to the xylem. This idea is compatible with both current models of strigolactone-dependent branching control²⁷.

According to the 'second messenger model', strigolactones are transported to the bud as a second messenger of auxin²⁸, hence cellular transport of strigolactones in the axillary regions would be indispensable. In the 'auxin transport canalization-based model', strigolactones are thought to dampen polar auxin transport, resulting in the accumulation of auxin to levels that inhibit bud outgrowth²⁹. Strigolactones could restrict auxin transport systemically and/or locally²⁷. For both models, local strigolactone transport capacity near the axils would be in line with the inhibitory role of strigolactones on branching. In *W115* × *W138-pdr1* plants, bud outgrowth was initiated sooner (Supplementary Fig. 8a) and more vigorously than in the wild-type plants, causing longer branches (Fig. 4e and Supplementary Fig. 8e–h) at node three to five. This early and vigorous bud outgrowth was also observed in *pdr1*-RNAi lines (Fig. 4f). *Dad1* mutants initiate branches from all nodes (Supplementary Fig. 8d), and although branch elongation is retarded compared with the wild type, these mutants eventually produce full branches from every node^{20,30}. At flowering time, this results in a phenotype that is more pronounced than in any of the *pdr1* mutants, which have final branch patterns that differ only marginally from the respective wild-type plants (Supplementary Fig. 8b, c and Supplementary Table 1). Nevertheless, the *pdr1* mutant branching phenotype seems to be strigolactone-dependent and related to the *dad1* mutant branching phenotype, because branch elongation in both mutants could be suppressed to near wild-type conditions by exogenous application of GR24 to the leaf axils (Fig. 4g, h).

In conclusion, the identification of PDR1-mediated strigolactone transport contributes to a more comprehensive view of the modes of action of strigolactones. Understanding the underlying transport mechanisms is crucial for a holistic view of phytohormone function. Strigolactone transport has a direct effect on the phosphate-dependent control of arbuscular mycorrhizal levels and on the control of shoot branching, for which the integration of auxin and strigolactone signalling seems to be achieved partially through reciprocal transport modulation. The mild branching phenotype of *pdr1* mutants relative to the strigolactone-biosynthesis mutant *dad1* (ref. 30) suggests that residual transport and/or locally produced strigolactones may compensate for defective strigolactone transport in the shoot. However, arbuscular mycorrhizal development was affected to a similar degree in the *pdr1* and *dad1* mutants, revealing that below-ground strigolactone transport and secretion relies primarily on PDR1.

METHODS SUMMARY

All experiments, with the exception of the transport assays, were performed in accordance with established protocols. Detailed methods and associated references can be found in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 November 2010; accepted 19 January 2012.

Published online 7 March 2012.

- Cook, C. E., Whichard, L. P., Turner, B., Wall, M. E. & Egley, G. H. Germination of witchweed (*Striga lutea* Lour.): isolation and properties of a potent stimulant. *Science* **154**, 1189–1190 (1966).
- Yoder, J. I. & Scholes, J. D. Host plant resistance to parasitic weeds: recent progress and bottlenecks. *Curr. Opin. Plant Biol.* **13**, 478–484 (2010).
- Akiyama, K., Matsuzaki, K.-I. & Hayashi, H. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* **435**, 824–827 (2005).
- Parniske, M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature Rev. Microbiol.* **6**, 763–775 (2008).
- Umeshara, M. et al. Inhibition of shoot branching by new terpenoid plant hormones. *Nature* **455**, 195–200 (2008).
- Gomez-Roldan, V. et al. Strigolactone inhibition of shoot branching. *Nature* **455**, 189–194 (2008).

- Petrasek, J. & Friml, J. Auxin transport routes in plant development. *Development* **136**, 2675–2688 (2009).
- Kuromori, T. et al. ABC transporter AtABCG25 is involved in abscisic acid transport and responses. *Proc. Natl. Acad. Sci. USA* **107**, 2361–2366 (2010).
- Kang, J. et al. PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proc. Natl. Acad. Sci. USA* **107**, 2355–2360 (2010).
- Matusova, R. et al. The strigolactone germination stimulants of the plant-parasitic *Striga* and *Orobanche* spp. are derived from the carotenoid pathway. *Plant Physiol.* **139**, 920–934 (2005).
- Verrier, P. J. et al. Plant ABC proteins — a unified nomenclature and updated inventory. *Trends Plant Sci.* **13**, 151–159 (2008).
- Moons, A. Transcriptional profiling of the *PDR* gene family in rice roots in response to plant growth regulators, redox perturbations and weak organic acid stresses. *Planta* **229**, 53–71 (2008).
- Badri, D. V. et al. An ABC transporter mutation alters root exudation of phytochemicals that provoke an overhaul of natural soil microbiota. *Plant Physiol.* **151**, 2006–2017 (2009).
- Sugiyama, A., Shitan, N. & Yazaki, K. Signaling from soybean roots to rhizobium: an ATP-binding cassette-type transporter mediates genistein secretion. *Plant Signal. Behav.* **3**, 38–40 (2008).
- Jasinski, M. et al. A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell* **13**, 1095–1107 (2001).
- Hayward, A., Stirnberg, P., Beveridge, C. & Leyser, O. Interactions between auxin and strigolactone in shoot branching control. *Plant Physiol.* **151**, 400–412 (2009).
- Hanlon, M. T. & Coen, C. Genetic evidence for auxin involvement in arbuscular mycorrhiza initiation. *New Phytol.* **189**, 701–709 (2011).
- Sharda, J. N. & Koide, R. T. Can hypodermal passage cell distribution limit root penetration by mycorrhizal fungi? *New Phytol.* **180**, 696–701 (2008).
- Koes, R. et al. Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. *Proc. Natl. Acad. Sci. USA* **92**, 8149–8153 (1995).
- Snowden, K. C. et al. The Decreased apical dominance1/*Petunia hybrida* CAROTENOID CLEAVAGE DIOXYGENASE8 gene affects branch production and plays a role in leaf senescence, root growth, and flower development. *Plant Cell* **17**, 746–759 (2005).
- Reddy, D. M. R. S., Schorderet, M., Feller, U. & Reinhardt, D. A petunia mutant affected in intracellular accommodation and morphogenesis of arbuscular mycorrhizal fungi. *Plant J.* **51**, 739–750 (2007).
- Zhang, Q., Blaylock, L. A. & Harrison, M. J. Two *Medicago truncatula* half-ABC transporters are essential for arbuscule development in arbuscular mycorrhizal symbiosis. *Plant Cell* **22**, 1483–1497 (2010).
- Kohlen, W. et al. Strigolactones are transported through the xylem and play a key role in shoot architectural response to phosphate deficiency in nonarbuscular mycorrhizal host *Arabidopsis*. *Plant Physiol.* **155**, 974–987 (2011).
- Ruyter-Spira, C. et al. Physiological effects of the synthetic strigolactone analog GR24 on root system architecture in *Arabidopsis*: another belowground role for strigolactones? *Plant Physiol.* **155**, 721–734 (2011).
- Beveridge, C. A. & Kyozuka, J. New genes in the strigolactone-related shoot branching pathway. *Curr. Opin. Plant Biol.* **13**, 34–39 (2010).
- Mashiguchi, K. et al. Feedback-regulation of strigolactone biosynthetic genes and strigolactone-regulated genes in *Arabidopsis*. *Biosci. Biotechnol. Biochem.* **73**, 2460–2465 (2009).
- Domagalska, M. A. & Leyser, O. Signal integration in the control of shoot branching. *Nature Rev. Mol. Cell Biol.* **12**, 211–221 (2011).
- Brewer, P. B., Dun, E. A., Ferguson, B. J., Rameau, C. & Beveridge, C. A. Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and *Arabidopsis*. *Plant Physiol.* **150**, 482–493 (2009).
- Crawford, S. et al. Strigolactones enhance competition between shoot branches by dampening auxin transport. *Development* **137**, 2905–2913 (2010).
- Napoli, C. highly branched phenotype of the petunia *dad1-1* mutant is reversed by grafting. *Plant Physiol.* **111**, 27–37 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We kindly thank the following: C. Gübeli for technical assistance; and T. Gerats, S. Hörtsteiner, A. Osbourne, P. Schläpfer and C. Beveridge for comments. This study was funded by the Swiss National Foundation within the NCCR-Plant Survival (project 'ABC transporters involved in signaling') and by The Netherlands Organization for Scientific Research (NWO; VICI grant 865.06.002 and Equipment grant 834.08.001 to H.J.B.). H.J.B. was co-financed by the Centre for BioSystems Genomics (CBSG).

Author Contributions T.K. wrote the manuscript, designed the project and carried out most of the experiments. W.K. and R.B. carried out the *P. hybrida* strigolactone analysis and the *P. ramosa* bioassays. J.S. performed the quantitative PCR with reverse transcription assays and the transport assays. J.S. and M.S. performed branching and GUS trials. L.B. analysed the *PDR1*-OE lines. J.B.B. sectioned material. D.R. investigated arbuscular mycorrhizal morphology. H.J.B. supervised the analytical part of the project. E.M. conceived and supervised the project. D.R., E.M., J.S., W.K. and H.J.B. assisted in editing the manuscript.

Author Information Sequences for *P. hybrida* *PDR1*, a 1.8-kb element upstream of *PDR1* and the *P. axillaris* orthologue of *PDR1* have been deposited in the GenBank database under accession numbers JQ292813, JQ292814 and JQ292812. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.K. (t.kretzschmar@irri.org).

METHODS

Plant growth conditions. Petunia lines were grown with a 16 h light cycle, at 60% relative humidity and 25 °C in soil (ED 73, Einheitserde) or in clay granules (Oil Dri US-Special, Damolin). Clay granules were supplemented once a week with half-strength Hoagland solution. For mycorrhization trials, a mix of 40% (v/v) soil, 40% (v/v) clay granules, 10% (v/v) sand and 10% (v/v) mycorrhizal inoculum (Agrauxine) was used. Seeds were plated on medium containing 2.2 g l⁻¹ MS medium (Duchefa) and 15 g l⁻¹ sucrose, supplemented with 9 g l⁻¹ Phyto Agar (Duchefa), with a 16 h light cycle at 25 °C. *Arabidopsis* was grown in vertical plates on medium containing 2.2 g l⁻¹ MS medium and 15 g l⁻¹ sucrose, supplemented with 9 g l⁻¹ Phyto Agar, with a 16 h light cycle and 60% relative humidity at 21 °C. For hormone treatment, 14-day-old W115 seedlings grown on plates were exposed for 24 h to final concentrations of 1 μM or 10 μM of the synthetic strigolactone analogue GR24 (Chiralix), 10 μM ABA or 25 μM NAA. For phosphate starvation, 14-day-old W115 seedlings were transferred to phosphate-free plates for 1 week.

PDR1 cloning strategy. *PDR1*-specific 0.5-kb transcripts were amplified from W115 root cDNA 5 weeks post inoculation (w.p.i.) with *G. intraradices*, using the following primers: 5'-mgwtagctctdyktkkgaccc-3' and 5'-gytctytytgncchc chgaaatwcc-3' (5' region), or 5'-gggwaaracggwgtiyagttggwgcw-3' and 5'-ctcatnaca atdgcwgwgcctwg-3' (3' region). Sequences follow the universal degenerate code. The resultant 5' and 3' fragments were aligned, and the deduced consensus primers 5'-TATTGGGACTTGAAATTGTGCCGATAC-3' and 5'-GCTCCACTAACAC CCATCAGAGCTGTC-3' were used to amplify putative *PDR1* fragments spanning 2.5 kb. The 5' and 3' ends of *P. hybrida* *PDR1* were amplified using a SMART RACE Amplification Kit (Clontech) with 5' RACE (5'-CTCGAGTACATTCTCGGCGGGACCTTGG-3'), nested 5' RACE (5'-CCATTTCTGTCCTCAACATG GTATCGG-3'), 3' RACE (5'-GTCCTCAAGAGTAGGAAGCATCTGCG-3') and nested 3' RACE (5'-ACCGAGGACCGGCTTGAACCTCTGAGAG-3'). The full-length *P. hybrida* *PDR1* cDNA GenBank accession number is JQ292813.

To obtain the full-length genomic sequence of *PDR1*, a *P. axillaris* BAC library (from C. Kuhlemeier) was screened with 5'-TGCAATCCTCATGATGTCA GTGG-3' and 5'-CCTCTCTCTCCTAGACAGCTCTGC-3'. *P. axillaris* is a progenitor of the hybrid species *P. hybrida*³¹, and a genomic library for BAC-based cloning was available only for *P. axillaris*. BACs were extracted from candidate clones via the Large-Construct Kit (QIAGEN). Full-length genomic *P. axillaris* *PDR1* was amplified from a BAC with 5'-AATTACTAGTATGGAGGGTG GTGAAG-3' and 5'-AATTGCATGCCTATCTTCTGAAATTAAATG-3', cut with SpeI and SphI, and cloned into the vector pUC18-GFP5Tsp through compatible NheI and SphI restriction sites, for use in GFP localization studies³². For stable transformation, the CaMV 35S::GFP::gPaPDR1-terminator cassette was cloned from pUC18-GFP5Tsp into the pGreenII 0179 vector³³ by using the following strategy. First, the CaMV 35S promoter from native pUC18-GFP5Tsp was cloned via XhoI and XmaI sites. Second, the terminator from the native pUC18-GFP5Tsp, including the upstream SphI site, was cleaved with NheI and SacI and inserted into CaMV 35S-pGreenII 0179 via compatible XbaI and SacI sites. Third, GFP::gPaPDR1 was cut and cloned via XmaI and SphI. Nine kilobases of the genomic *P. axillaris* *PDR1* locus, including upstream and downstream sequence, were submitted to GenBank under accession number JQ292812.

PDR1 promoter GUS construct and GUS staining assay. A 1.8-kb *P. hybrida* *PDR1* promoter fragment was amplified using a GenomeWalker Universal Kit (Clontech) with 5'-AGTGGAAAGTTCTCAAGTCAGCCCCA-3' and the nested 5'-CCCTAAAGAGTCITCACCACCCCTCAT-3' (GenBank accession number JQ292814). The fragment was cloned into the pGEM-T-Easy Vector system (Promega), reamplified with 5'-CATGAAGCTTGACCCAGAAGAAG ATTAGGC-3' and 5'-TCGATCTAGACACATTAAGAGGAAAGTAGGTAC-3' and cloned into the pGPTV-BAR³⁴ vector system via HindIII and XbaI. Of the original T0 transformants, eight lines were selected for further analysis. Segregating T1 individuals of all eight lines had comparable below-ground and above-ground expression patterns at different developmental stages. Two of these lines were chosen for the in-depth analysis presented in this work, and all data presented were confirmed in both lines.

GUS-staining trials were performed as described previously³⁵. After staining, samples were cleared for 24 h in 10% (w/v) KOH and stored in 70% (v/v) ethanol. For the analysis of HPCs, 5-bromo-6-chloro-3-indolyl β-D-glucuronide cyclohexylammonium salt was used for GUS staining, and samples were cleared for 24 h in 10% (w/v) KOH before staining with trypan blue as described previously¹⁸.

PDR1 RNAi constructs. Silencing of *P. hybrida* *PDR1*-specific transcripts was performed with the pKANNIBAL vector system³⁶. Two constructs were designed: one targeting a highly variable region within nucleotide binding domain 2 (NBD2) of *PDR1* (the C construct), and the other targeting part of the 3' end and the 3' untranslated region of *PDR1* (the R construct). The 148-base-pair (bp) C fragment (5'-GGAACGCAAGCAAAGGGGTGAGGTTATTGAACATCTCGCTTGG AAAGAGCTCTGAAAAGGAAATGATGTTCGCGAAGTGCATCTTC

CAGGTCAATGTCCTCAAGAGTAGGAAGCATCACTGCGGCTGATTGAG CAAGAG-3') was amplified from W115 cDNA with 5'-CGATGGATCCTCGA GGGAACGCAAGCAAAGGG-3', containing BamHI and XhoI restriction sites and with 5'-CGATATCGATGGTACCCCTCTGCTCAAATCAGCCGAGTGA-3' containing ClaI and KpnI sites. The 411-bp R fragment (5'-GACATTATATG GACTAATTGCTCACAATTGAGACATACAAGACAGACTTGACACAA ATGAGACAGTGGAAACAATTCTAGAGAATTCTTGTATTCAAACATG ATTGTTGAGGATATGTTGCTCTCATTTCTGGGATTCTGTCTTCTTCTCTCATTTGCTATTCAATTAAAACATTAAATTCCAGAAAAGA TAGGTTGGTCCAGGTACACATGAAAAGAGCGTTATCAAGATATGT GTATATTAGGATAATAATATAATCTTTCTCTCTTACTTATT GTGGTTTCTCAAGTTGGAATAGATAGAACCAAAGTCTGACTCTG TATTTAAGAACAACTTTGTACACATTGTTATGTATTGGAGAAGTTATG AGTATCTTTG-3') was amplified with 5'-CGATGGATCCTCGAGACATTA TATGGACTAATTGCC-3' containing BamHI and XhoI restriction sites and with 5'-CGATATCGATGGTACCAAAAGATACTCATAACTCTCC-3' containing ClaI and KpnI sites. The resultant amplicons were cloned in the sense and antisense direction in the two multiple cloning sites of pKANNIBAL. The pKANNIBAL RNAi cassette was excised from the vector backbone via NotI and transferred into the binary pGreenII 0229 vector³³. After stable transformation of W115, the extent of downregulation was estimated by semiquantitative PCR or PCR with reverse transcription.

Plant transformation. W115 was transformed as described previously³⁷. Construct insertion was confirmed by PCR on genomic DNA with the primers 5'-ACGGTCCACATGCCGGTATATACGATG-3' and 5'-GATGGCATTGTA GGAGCCACCTCC-3', which target the CaMV 35S promoter, or with 5'-GAA TTGATCAGCGTTGGTGGGAAAGC-3' and 5'-GGTAATGCGAGGTACGGT AGGAGTTG-3', which target the GUS gene.

Transient transformation of *Arabidopsis* Col-0 protoplasts was performed as described previously³². *Arabidopsis* plants were stably transformed as described previously³⁸. The T0 generation was selected for hygromycin resistance. Plants of the T1 generation were tested for hygromycin resistance and GFP expression.

Screening approach to identify transposon insertions in PDR1. A 3D-gDNA library (from T. Gerats) representing 10 × 10 × 10 (1,000) W138 individuals was screened for dTph1 insertions in *P. hybrida* *PDR1* using a PCR-based method³⁹. The entire genomic region of *PDR1* was scanned in contiguous steps covering less than 1 kb, using the dTph1-specific primer 5'-GAATTGCTCCGCCCTG-3' and a variety of ³³P-labelled gene-specific primers. The primer 5'-CCATTTGTC TCCAACAATGGTATCGG-3' yielded a positive result. Heterozygosity PCRs were performed with transposon-flanking primers: 5'-tgccaatccctatgtatgtcagg-3' and 5'-ccttcctctctcgtacagtcgttc-3'. Homozygous dTph1 insertion alleles were crossed into W115, the progeny were selfed, and the resultant offspring were tested for homozygosity with the abovementioned primers. Transposon display analysis using six W115 × W138 and five W115 × W138-pdr1 lines was performed as described previously⁴⁰.

Mycorrhization trials. Subsets of mycorrhized roots were stained⁴¹ and quantified for their level of colonization using the gridline intersect method⁴². Only roots with clear intraradical structures such as coiled cortical hyphae, arbuscules and vesicles were scored as positively mycorrhized. A minimum of 200 intersecting root fragments per sample were investigated microscopically for intraradical arbuscular mycorrhizal structures. For trials involving W115 × W138 lines, five individuals from five lines were analysed, and the data were presented as a pool of $n = 25$ (5 × 5). Double staining of the colonized roots with propidium iodide and wheat germ agglutinin coupled to fluorescein isothiocyanate (WGA-FITC, Sigma-Aldrich) was performed as described previously⁴³.

Hyphal branching bioassays. Branching assays were performed as described previously⁴⁴ with preselected spores of *G. margarita* (Agrauxine). For the production of the root exudate concentrate, petunia lines were grown in clay granules and then transferred for 24 h to 0.1 l of a hydroponic solution containing 2 mM CaCl₂ and 2 mM KSO₄ and kept under constant aeration. The hydroponic solution was then passed through a Sep-Pak Classic C18 Cartridge (Waters) to adsorb hydrophobic root exudates. Exudates were eluted from the column using 2 ml of acetone, and the eluent was dried over nitrogen. Dried exudates were redissolved in acetone and normalized according to root fresh weight. Exudate equivalents of 10 mg root fresh weight were used in each branching assay.

Transport assays and GR24 tolerance assays. *Arabidopsis* seeds from three independent *PDR1*-OE lines were surface-sterilized with 1% (v/v) bleach and 50% (v/v) ethanol and plated on media supplemented with 2.2 g l⁻¹ MS medium, 1% (w/v) sucrose and 0, 10 or 25 μM GR24. After three days of stratification, plants were moved to a 16 h light cycle and selected for GFP fluorescence after 3 days of growth. Root length was determined with ImageJ 1.44 software (<http://rsbweb.nih.gov/ij>) after 7 days, and seedlings were moved to hygromycin-containing plates without sucrose to confirm the selection by GFP fluorescence.

For transport experiments, seeds were sterilized in the same way and plated on hygromycin-containing medium. After 3 days of stratification and 3 days of growth, seedlings were checked for GFP fluorescence. After 7 days, GFP- and hygromycin-positive plants were transferred to medium supplemented with 2.2 g l^{-1} MS medium and 1% (w/v) sucrose and grown for another 7 days. Three *PDR1*-OE lines, Col-0 and a vector control line were incubated for 2 h in the dark at 4°C with root tips submerged in 0.1% (w/v) Phyto Agar supplemented with 25 nM [^3H]GR24 (specific activity 40 Ci mmol $^{-1}$; American Radiolabeled Chemicals). Subsequently, the plant roots were washed in ice-cold 1 mM CaCl_2 and incubated in 200 μl 0.1% Phyto Agar. For each line, 50% of the plants were kept for a further 1 h at 4°C in the dark as a diffusion control, and the other 50% were shifted to 23°C for 1 h to monitor transport. Subsequently, the shoot, root and Phyto Agar fractions were incubated for 30 min in 50 μl 24% (w/v) trichloroacetic acid at 23°C . ^3H counts were determined in 3 ml ULTIMA Gold LSC cocktail (Perkin Elmer) with a Liquid Scintillation Analyser Tri-Carb 2900TR (Packard BioScience). The number of disintegrations per minute was computed into a percentage for each fraction and normalized to tissue fresh weight.

RNA isolation, cDNA synthesis, semiquantitative PCR and quantitative RT-PCR. RNA was isolated with an RNeasy Plant Mini Kit (QIAGEN). Reverse transcription of RNA to cDNA was performed with M-MLV Reverse Transcriptase (Promega) and a poly(T) primer (Promega).

P. hybrida PDR1 expression was quantified semiquantitatively with the primers 5'-GAAACTGTGGCCGAAAGG-3' and 5'-GAGTTCAAGCCGGTCT-3' or 5'-AAATGCTACTACAGTGCAG-3' and 5'-CATATAATGTCCAGGAAATGG-3'. Tubulin 1 (TUB) transcripts, partially amplified with 5'-CATTG GTCAAGCCGGTTATTC-3' and 5'-ACCCCTTGAAGACCAGTACAGT-3', served as a housekeeping and loading control.

Samples were diluted 1:30, and the diluted samples (4 μl) were added to each reaction well, serving as template for the reaction. Deionized water served as a negative control for amplification. *PDR1* expression was quantified with the primers 5'-CCTGAGGTTTACCAATGGG-3' and 5'-GATGGTATTGGATTGGAGCA-3'. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) expression was quantified with the primers 5'-GACTGGAGAGGTGGAAGAGC-3' and 5'-CCGTTAAG AGCTGGGAGAAC-3'. GAPDH served as a housekeeping gene for normalization because it has been shown not to be regulated by hormonal treatments or mycorrhization (D. Reinhardt, personal communication). Final primer concentrations of 50, 100, 200 and 300 nM were tested for cDNA amplification and melting behaviour in the range of 60°C to 95°C . Because no differences were recorded, the average concentration of 100 nM was chosen for further experiments. Primer efficiency was recorded with W115 root cDNA as a template in a dilution range of 1:1 to 1:512, resulting in 94.42% for *P. hybrida PDR1* and 98.561% for GAPDH. These values were taken into account in the calculations. SYBR Green PCR Master Mix (Applied Biosystems) was added to the samples to a 20 μl final volume. For each sample, three technical replicates were pipetted. RT-PCR was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems) with 7500 Software version 2.0.4. For quantification, comparative C_T ($\Delta\Delta C_T$) was chosen as the method 45 ; the PCR run was divided into three parts: first, the hold stage (50°C for 2 min and 95°C for 10 min); second, the cycling stage (95°C for 15 s and 60°C for 1 min for 40 cycles); and third, the melt curve stage (95°C for 15 s, 60°C to 95°C over 1 min, 95°C for 30 s, and 60°C for 15 s). Relative differences were calculated as described previously 45,46 . Each experiment was performed for three biological replicates.

Isolation, identification and quantification of petunia strigolactones. Plants were grown in an X-Stream 20 aeroponic system (Nutriculture) as previously described for *Medicago truncatula* 47 . From day 8 until day 12, exudates were collected and pooled, and root material was sampled. Samples were stored at -80°C until analysis. *P. hybrida* root exudates and extracts were prepared and analysed by ultra performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) as previously described for *Arabidopsis* 22 . Orobanchol was provided by K. Yoneyama. For trials involving W115 \times W138 lines, three individuals from three lines were analysed, and the data are presented as a pool of $n = 9$ (3×3).

P. ramosa germination bioassay. Germination assays with *P. ramosa* seeds were conducted as reported previously 1 . Exudates were prepared as described above (in the subsection 'Hyphal branching bioassays'). GR24 (0.1 or 1.0 nM) and demineralised water were included as positive and negative controls, respectively. *P. ramosa* seeds were provided by M. Vurro. For trials involving W115 \times W138 lines, two individuals from five lines were analysed, and the data are presented as a pool of $n = 10$ (5×2).

Trypan blue staining of HPCs. Trypan blue staining of HPCs in roots was performed as described previously 18 .

Axillary branching trials. For a comparative analysis of lateral branch production in wild-type and *pdr1* backgrounds, plants were grown for 65 days in 0.55 l pots in soil as described above (in the subsection 'Plant growth conditions') and were watered daily. Branch development was monitored at different time points in a binomial fashion (yes/no) with respect to the following parameters: bud length >7 mm and full branch. Full branches were scored in accordance to a petunia branch definition 48 . Furthermore, branch length was scored as a continuous parameter. For trials involving W115 \times W138, 22 individuals from 5 lines were analysed, and the data are presented as a pool of $n = 110$ (5×22). For branch length trials in response to GR24 treatments, three lines each of W115 \times W138, W115 \times W138-*pdr1*, V26 and the *dad1* mutant (from K. Snowden) were grown in soil as described above (in the subsection 'Plant growth conditions'). From 25 to 40 days post germination, plants were treated three times each week with 0 μM or 10 μM GR24 as described previously 6 . For trials involving W115 \times W138 lines, eight individuals from three lines were analysed, and the data are presented as a pool of $n = 24$ (3×8).

Statistical analyses. Depending on the experimental set-ups and prerequisites, Students *t*-tests, Fisher's Exact tests or generalized linear models (GLMs) with quasi-binomial error structures were applied using R software.

Bioinformatics. Analysis of DNA fragments and vector constructs was performed using Vector NTI software (Invitrogen). The membrane topology of *PDR1* was predicted using ConPredII 49 . Phylogenetic analysis of *PDR1* was performed using tools available at Phylogeny.fr (<http://www.phylogeny.fr>). Alignments were performed with MultAlin software (<http://multalin.toulouse.inra.fr/multalin/>).

Robustness of data sets. All data sets presented were confirmed in at least two independent trials with similar set-ups and outcomes. For mycorrhization and branching trials, individual pots were randomized to reduce positional effects, and sample size was kept high to reduce background effects.

31. Zhang, X., Nakamura, I. & Mii, M. Molecular evidence for progenitorial species of garden petunias using polymerase chain reaction-restriction fragment length polymorphism analysis of the *Chs-1* gene. *HortScience* **43**, 300–303 (2008).
32. Meyer, A., Eskandari, S., Grallath, S. & Rentsch, D. AtGAT1, a high affinity transporter for γ -aminobutyric acid in *Arabidopsis thaliana*. *J. Biol. Chem.* **281**, 7197–7204 (2006).
33. Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S. & Mullineaux, P. M. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* **42**, 819–832 (2000).
34. Becker, D., Kemper, E., Schell, J. & Masterson, R. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol. Biol.* **20**, 1195–1197 (1992).
35. Cervera, M. Histochemical and fluorometric assays for *uidA* (GUS) gene detection. *Methods Mol. Biol.* **286**, 203–213 (2004).
36. Wesley, S. V. *et al.* Construct design for efficient, effective and high-throughput gene silencing in plants. *Plant J.* **27**, 581–590 (2001).
37. Lutke, W. K. Petunia (*Petunia hybrida*). *Methods Mol. Biol.* **344**, 339–349 (2006).
38. Harrison, S. *et al.* A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation. *Plant Methods* **2**, 19 (2006).
39. Vandenbussche, M. & Gerats, T. TE-based mutagenesis systems in plants: a gene family approach. *Methods Mol. Biol.* **260**, 115–127 (2004).
40. Van den Broeck, D. *et al.* Transposon display identifies individual transposable elements in high copy number lines. *Plant J.* **13**, 121–129 (1998).
41. Vierheilig, H., Coughlan, A., Wyss, U. & Piche, Y. Ink and vinegar, a simple staining technique for arbuscular-mycorrhizal fungi. *Appl. Environ. Microbiol.* **64**, 5004–5007 (1998).
42. Giovannetti, M. & Mosse, B. An evaluation of techniques for measuring vesicular arbuscular mycorrhizal infection in roots. *New Phytol.* **84**, 489–500 (1980).
43. Feddermann, N. *et al.* The *PAM1* gene of petunia, required for intracellular accommodation and morphogenesis of arbuscular mycorrhizal fungi, encodes a homologue of VAPYRIN. *Plant J.* **64**, 470–481 (2010).
44. Nagahashi, G. & Douds, D. D. Rapid and sensitive bioassay to study signals between root exudates and arbuscular mycorrhizal fungi. *Biotechnol. Techniques* **13**, 893–897 (1999).
45. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
46. Yuan, J. S., Reed, A., Chen, F. & Stewart, C. N. Statistical analysis of real-time PCR data. *BMC Bioinformatics* **7**, 85 (2006).
47. Liu, W. *et al.* Strigolactone biosynthesis in *Medicago truncatula* requires the symbiotic GRAS-TYPE transcription factors NSP1 and NSP2. *Plant Cell* **23**, 3853–3865 (2011).
48. Snowden, K. C. & Napoli, C. A. A quantitative study of lateral branching in petunia. *Funct. Plant Biol.* **30**, 987–994 (2003).
49. Arai, M. *et al.* ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic Acids Res.* **32**, 390–393 (2004).

Goblet cells deliver luminal antigen to CD103⁺ dendritic cells in the small intestine

Jeremiah R. McDole¹*, Leroy W. Wheeler²*, Keely G. McDonald², Baomei Wang¹, Vjollca Konjufca³, Kathryn A. Knoop², Rodney D. Newberry² & Mark J. Miller¹

The intestinal immune system is exposed to a mixture of foreign antigens from diet, commensal flora and potential pathogens. Understanding how pathogen-specific immunity is elicited while avoiding inappropriate responses to the background of innocuous antigens is essential for understanding and treating intestinal infections and inflammatory diseases. The ingestion of protein antigen can induce oral tolerance, which is mediated in part by a subset of intestinal dendritic cells (DCs) that promote the development of regulatory T cells¹. The lamina propria (LP) underlies the expansive single-cell absorptive villous epithelium and contains a large population of DCs (CD11c⁺ CD11b⁺ MHCII⁺ cells) comprised of two predominant subsets: CD103⁺ CX₃CR1⁻ DCs, which promote IgA production, imprint gut homing on lymphocytes and induce the development of regulatory T cells²⁻⁹, and CD103⁻ CX₃CR1⁺ DCs (with features of macrophages), which promote tumour necrosis factor- α (TNF- α) production, colitis, and the development of T_H17 T cells^{5-7,10}. However, the mechanisms by which different intestinal LP-DC subsets capture luminal antigens *in vivo* remains largely unexplored. Using a minimally disruptive *in vivo* imaging approach we show that in the steady state, small intestine goblet cells (GCs) function as passages delivering low molecular weight soluble antigens from the intestinal lumen to underlying CD103⁺ LP-DCs. The preferential delivery of antigens to DCs with tolerogenic properties implies a key role for this GC function in intestinal immune homeostasis.

We examined the *in vivo* antigen acquisition behaviour of intestinal LP-DCs in fluorescent DC-reporter mice using two-photon microscopy (Supplementary Fig. 1a). The intestine is imaged within the peritoneal cavity and images can be acquired from either the intact intestinal serosa or from the luminal surface through a small longitudinal incision in the intestine (Fig. 1a, b and Supplementary Movie 1, upper panels). The preparation is sufficiently stable to permit three-dimensional imaging of DC behaviour deep within intestinal tissues (Fig. 1c and Supplementary Movie 1, lower left panel) and preserves blood flow and epithelial barrier integrity for more than 4 h of continuous imaging.

We assessed antigen distribution by two-photon microscopy following the intraluminal injection of 10 kDa rhodamine dextran as a model antigen. Dextran coated the surface of the epithelium and filled the space between villi and crypts (Fig. 1a, b). In addition, we observed cylindrical dextran columns approximately 5 μ m in diameter and about 20 μ m long projecting through the villus epithelium and into the LP, when imaging from either serosal or luminal orientations (Fig. 1a-c, Supplementary Fig. 1b and Supplementary Movie 1, upper panels and lower right panel). Transepithelial dextran columns were common throughout the small intestine from duodenum to ileum, but did not cause a general disruption of the epithelial barrier as shown by the exclusion of dextran from the LP (Fig. 1a-c). We did not detect transepithelial dextran columns in the stomach, caecum, or colon, with the exception of the epithelium overlying the caecal patches

(Supplementary Fig. 1c-e and Supplementary Movie 2). Confocal microscopy revealed that dextran columns were a subset of epithelial cells containing intracellular dextran, which had a continuous border of e-cadherin on their basolateral surface and were often in contact with yellow fluorescent protein (YFP⁺) cells in the LP of CD11c-YFP reporter mice³⁰ (Fig. 1d, e and Supplementary Movie 4, right panel).

Periodic acid-Schiff staining of mucin in sections of the small intestine (Fig. 2a) produced a goblet cell (GC) staining pattern similar in frequency, distribution and dimensions to the dextran columns identified by two-photon microscopy (Fig. 2b, c). Furthermore, in contrast to the acellular and impermeable discontinuities seen in the small intestine epithelium¹¹, dextran columns were associated with a nucleus (Fig. 2d, Supplementary Fig. 2a and Supplementary Movie 3). To determine if the dextran-filled cells were in fact GCs, sections of intestine from mice given lysine-fixable dextran were stained with antibodies to mucin 2 (MUC2) and cytokeratin 18, which are both highly expressed by GCs¹². Dextran columns showed near perfect colocalization with MUC2⁺ and cytokeratin 18⁺ epithelial cells displaying GC morphology (Fig. 2e, f). Therefore, we term this phenomenon 'goblet-cell-associated antigen passages' (GAPs). To address the possibility that GAPs are apoptotic GCs, we co-stained for various markers of apoptosis, including cleaved cytokeratin 18, cleaved caspase 3 and TdT-mediated dUTP nick end labelling (TUNEL; Supplementary Fig. 3a-i). In all cases, we found no association between apoptotic GCs and GAPs. Moreover, GAPs are distinct from villous M cells, because they did not colocalize with the M-cell marker glycoprotein 2 (GP2) (Fig. 2g)¹³. The frequency and distribution of GAPs assessed by two-photon microscopy was similar in all strains of specific-pathogen-free (SPF) mice examined (Supplementary Fig. 2b-d), with a non-significant trend toward more GAPs detected in the terminal ileum (Supplementary Fig. 2h). GAPs were also evident in human jejunum resection specimens (Fig. 2h, i), suggesting that GAPs are a general phenomenon of the healthy small intestine. We examined the frequency of GAPs in C3H/HeJ/Bir IL-10^{-/-} mice¹⁴, which develop spontaneous intestinal inflammation with GC loss, and in germ-free mice that lack normal gut flora. The number of GAPs and GCs correlated strongly; GAPs and GCs were significantly more numerous in germ-free mice (Supplementary Fig. 2e, g) and significantly fewer in IL-10^{-/-} mice (Supplementary Fig. 2f, g).

Previous studies have shown that LP-DCs can extend transepithelial dendrites (TEDs) between intestinal epithelial cells to sample luminal contents and microbiota¹⁵⁻¹⁸. However, these studies used *ex vivo* and exteriorized tissue preparations that involved removing the luminal contents and mucous before imaging. Using our *in vivo* imaging preparation and confocal microscopy, we found that although LP-DCs probed the epithelium actively with their dendrites (Fig. 1c and Supplementary Movie 1, lower left panel, and Supplementary Movie 4), they did not extend TEDs into the intestinal lumen to capture fluorescent antigen in healthy mice (based on over 50 independent intravital imaging experiments examining all regions of the small intestine from the tip of the

¹Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²Department of Internal Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Department of Microbiology, Southern Illinois University, Carbondale, Illinois 62901, USA.

*These authors contributed equally to this work.

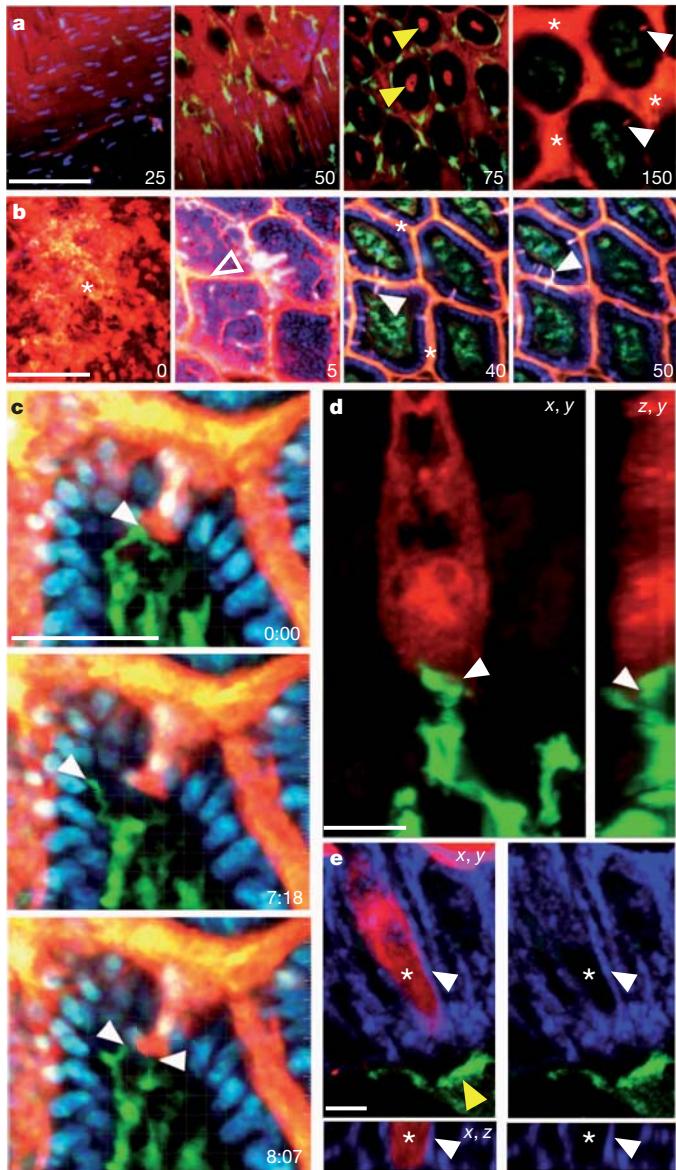


Figure 1 | Steady-state transepithelial delivery of luminal material in the mouse small intestine. **a, b**, Intestines from anaesthetized CD11c-YFP reporter mice injected intraluminally with 10 kDa dextran (red) imaged with intravitral two-photon microscopy from the (a) serosal or (b) luminal surfaces. Optical sections at increasing depths (shown in μm) revealed dextran in the lumen (asterisks), the crypts (yellow arrows), on the epithelial surface (unfilled arrow) and columns of dextran that traversed the epithelium (white arrows). Dextran was generally excluded from the LP as identified by CD11c-YFP⁺ LP-DCs (green) below the DAPI-stained epithelial nuclei (blue in b). Scale bars, 100 μm . **c**, Time-lapse recording of an LP-DC dendrite (green, white arrow) making repeated contacts with a dextran column (red) crossing the epithelium (DAPI-stained nuclei, blue). Scale bar, 50 μm ; time stamp shows min:s elapsed time from the start of imaging. **d**, Rendered confocal image of a CD11c-YFP⁺ DC (green) in contact with a dextran-filled epithelial cell (red). Panels show orthogonal views, contact is indicated by the white arrow. Scale bar, 5 μm . **e**, Confocal image of a dextran-containing cell (red) bordered by a continuous e-cadherin-positive (blue) surface (white arrow). Asterisk indicates the position of an optical section near the cell's centre showing intracellular dextran. A CD11c-YFP⁺ DC (green, yellow arrow) is positioned near the base of the epithelium. Orthogonal projection (bottom panels); red channel removed (right panels). Scale bar, 5 μm .

villi to the base of the crypts). To confirm that our imaging approach can readily detect TEDs, we imaged CD11c-YFP reporter mice challenged with *Salmonella typhimurium*, which has been shown to upregulate TED formation¹⁵. Although rare LP-DC TEDs were observed in approximately 2% of villi, TEDs did not mediate the uptake

of luminal dextran or beads on the basis of fluorescence colocalization (Supplementary Fig. 4a and Supplementary Movie 5, upper left panel).

Next, we examined whether paracellular leak¹⁹ could serve as a major source of luminal antigen for small intestine LP-DCs. The intraluminal injection of 5 mg of 10 kDa dextran produced a faint 'feather'-like staining pattern between villous epithelial cells, consistent with paracellular leak (Supplementary Fig. 4b and Supplementary Movie 5, upper right panel). Time-lapse two-photon imaging showed that dextran collected at the base of the epithelium, but was flushed out of the villi efficiently during contraction and did not remain colocalized with LP-DCs even in areas of extensive paracellular leak (Supplementary Fig. 4b and Supplementary Movie 4, left panel, and Supplementary Movie 5, upper right panel). In addition, we did not detect paracellular leak around GAPs in fixed sections by confocal microscopy, despite the more permissive tight junctions of GCs²⁰. However, we cannot exclude the possibility that LP-DCs capture low levels of antigen via paracellular leak, because this process might be below the level of detection of our imaging approach.

In contrast to TEDs and paracellular leak, two-photon time-lapse imaging provided direct evidence that GAPs are a source of luminal antigen for LP-DCs. In addition to dextran (Fig. 3a), GAPs were capable of transporting protein antigens (Fig. 3b). Although most GAPs remained visible for the duration of our imaging experiments, they were a dynamic phenomenon (Supplementary Movie 5, lower panels). Moreover, the manner in which LP-DCs interacted with GAPs varied. In some cases, DCs made stable contacts and slowly collected antigen over several minutes (Fig. 3a and Supplementary Movie 5, lower left panel) whereas in others, DCs actively probed GAPs and captured clumps of antigen (Fig. 3b and Supplementary Movie 5, lower right panel). We assessed the molecular weight exclusion limit of GAPs and found that beads ranging from 0.02 to 1.0 μm in size did not enter GAPs (Supplementary Fig. 4c). In contrast, GAPs filled rapidly with 10 kDa dextran (Supplementary Fig. 4d) and dextran colocalized with CD11c-YFP⁺ LP-DCs 2 h after intraluminal injection (Fig. 3a and Supplementary Movie 5, lower left panel). GAPs also filled with larger dextrans (Supplementary Fig. 4d); however, capture by LP-DCs was markedly reduced with 70 kDa dextran and undetectable with 2,000 kDa dextran during our 4-h imaging window.

The small intestine LP contains two prominent DC populations: CX₃CR1⁻ CD103⁺ DCs with tolerogenic potential and CX₃CR1⁺ CD103⁺ LP-DCs that have features of macrophages and have been implicated in intestinal inflammation^{5–10}. We visualized LP-DC subsets *in vivo* using dual-reporter mice created by crossing CD11c-YFP mice with CX₃CR1-GFP knock-in mice (ref. 28) in which CD103⁻ and CD103⁺ LP-DCs can be distinguished by the presence or absence of CX₃CR1-GFP expression, respectively (Fig. 3c and Supplementary Fig. 5). We frequently observed CD11c-YFP⁺ CX₃CR1-GFP⁻ LP-DCs sampling GAPs by two-photon microscopy, but this behaviour was rare in CD11c-YFP⁺ CX₃CR1-GFP⁺ LP-DCs (Fig. 3c and Supplementary Fig. 6c); out of 50 total LP-DCs containing dextran, 49 were CD11c-YFP⁺ CX₃CR1-GFP⁻. Furthermore, when we directly observed antigen transfer from GAPs to DCs (20 out of 500 GAPs examined), GAPs delivered antigen exclusively to CD11c-YFP⁺ CX₃CR1-GFP⁻ LP-DCs (Fig. 3d). Immunofluorescence microscopy confirmed that the cells interacting with GCs were CD103⁺ CD11c⁺ LP-DCs and not B220⁺ plasmacytoid DCs (pDCs) (Supplementary Fig. 6a). Moreover, flow cytometry showed that luminal antigen was captured preferentially by CD103⁺ DCs at a proportion of roughly 10:1 over CD103⁻ DCs (Fig. 3e) and rarely colocalized with pDCs (Supplementary Fig. 6b), consistent with the sampling bias observed by two-photon imaging. Control experiments demonstrated that antigen uptake during cell isolation was negligible (Supplementary Fig. 7), thus the flow cytometry results measure the *in vivo* antigen acquisition capacity of different LP-DC subsets. Interestingly, CD103⁺ LP-DCs, but not CD103⁻ LP-DCs, often stained positive for cytokeratin 18 (Fig. 3f and g), which is highly expressed by villous GCs (Fig. 3h).

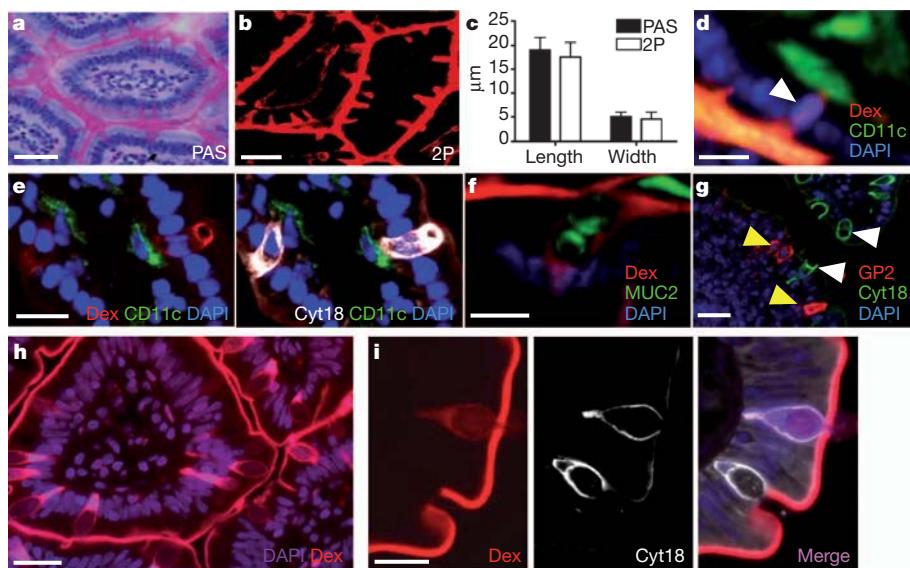


Figure 2 | GCs are associated with the transepithelial passage of luminal material. **a, b**, Periodic acid–Schiff (PAS) staining of GCs (**a**) and dextran columns (**b**) visualized by two-photon microscopy. **c**, PAS and dextran stained structures have similar morphology and dimensions. **d**, Dextran columns were often associated with a nucleus (white arrow). **e, f**, Dextran columns (Dex, red)

colocalized with GC markers cytokeratin 18 (Cyt18, white) (**e**) and MUC2 (**f**). **g**, Cytokeratin 18-positive cells (white arrows) did not colocalize with the M-cell marker GP2 (yellow arrows). **h, i**, Dextran columns were present in healthy human small intestine (**h**) and stained positive for cytokeratin 18 (**i**). Scale bars, 30 μ m (**a, b** and **h**), 20 μ m (**d, e, g** and **i**), and 10 μ m (**f**). Error bars, s.d.

Because cytokeratin 18 expression is undetectable in LP-DCs by quantitative real-time PCR (Fig. 3h), this suggests that CD103⁺ LP-DCs interact selectively with GAPs and can capture GC-derived proteins.

In two-photon imaging experiments, LP-DCs captured fluorescent luminal ovalbumin (Ova) readily from GAPs (Fig. 4a and Supplementary Movie 6), similarly to fluorescent dextran and BSA. To determine if LP-DCs could process and present luminal antigen, we administered Ova *in vivo*, then sorted LP-DCs and co-cultured them with OTI T cells²¹. T-cell proliferation was assessed on day 3 by both CFDA (dye) dilution (Fig. 4b, c) and by counting T cells after culture (Fig. 4d). Total LP-DC populations (CD45⁺, CD11c⁺, MHCII⁺) were capable of inducing modest OTI T-cell proliferation that was significantly greater than controls (Fig. 4b, d). CD103⁺ LP-DCs stimulated significant OTI T-cell proliferation, whereas CD103⁻ LP-DCs from the same mice did not (Fig. 4d). The failure of CD103⁻ LP-DCs to stimulate T cells was not due to a lack of intrinsic antigen presentation capacity, because they were capable of inducing comparable levels of OTI T-cell proliferation to CD103⁺ LP-DCs (81.4% to 88.2%, respectively) when Ova was added to the cell culture (Fig. 4c). These findings suggest that CD103⁺ LP-DCs have cross-presentation capacity, similar to their mesenteric lymph node counterparts².

In some two-photon time-lapse recordings we observed GAPs forming or disappearing (Supplementary Movies 3 and 5, lower panels), suggesting that GAP formation could be related to changes in GC function or secretion. To test whether GC secretion was associated with GAP

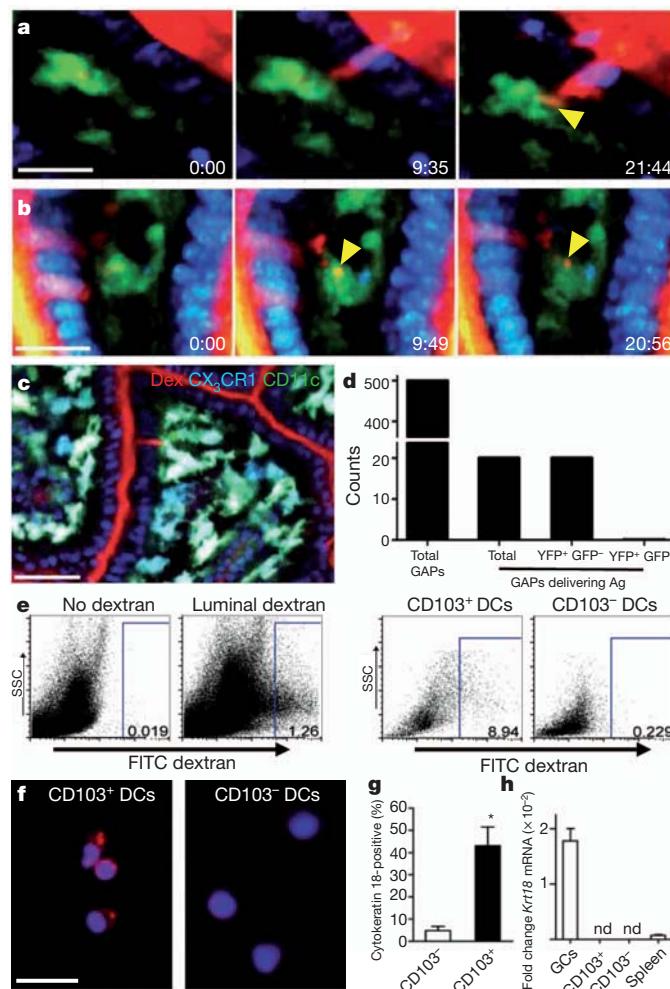


Figure 3 | GAPs deliver soluble antigen to CD103⁺ LP-DCs in the steady state. **a, b**, Time-lapse two-photon imaging of model antigens (red) dextran (**a**) and bovine serum albumin (BSA) (**b**) delivered by GAPs to LP-DCs (green, CD11c-YFP). Antigen from GAPs colocalized with LP-DCs (yellow arrows) over time. **c**, Two-photon imaging of CD11c-YFP CX₃CR1-GFP mice. GAPs (red) delivered antigen preferentially to CD103⁺ LP-DCs (CD11c-YFP⁺ CX₃CR1-GFP⁺; green) over CD103⁻ LP-DCs (CD11c-YFP⁺ CX₃CR1-GFP⁻; cyan) LP-DCs. **d**, Enumeration of GAPs and GAPs delivering antigen to LP-DC subtypes in CD11c-YFP CX₃CR1-GFP mice in two-photon recordings. **e**, Flow cytometry of LP cells showed that CD103⁺ DCs captured more luminal dextran than CD103⁻ DCs. FITC, fluorescein isothiocyanate. SSC, side scatter. **f**, Cytospins on sorted LP-DCs stained with DAPI (blue) and the GC marker cytokeratin 18 (red). **g**, Significantly more CD103⁺ LP-DCs than CD103⁻ LP-DCs stained cytokeratin 18-positive per high-powered field ($P = <0.001$). **h**, Neither DC population expressed detectable cytokeratin 18 mRNA. Scale bar, 15 μ m (**a** and **b**), 50 μ m (**c**), 25 μ m (**f**). Data in **g** taken from nine or more high-powered fields (representing more than 150 cells). Data in **h** performed in triplicate and is representative of two experiments. Error bars, s.e.m. nd, not detected.

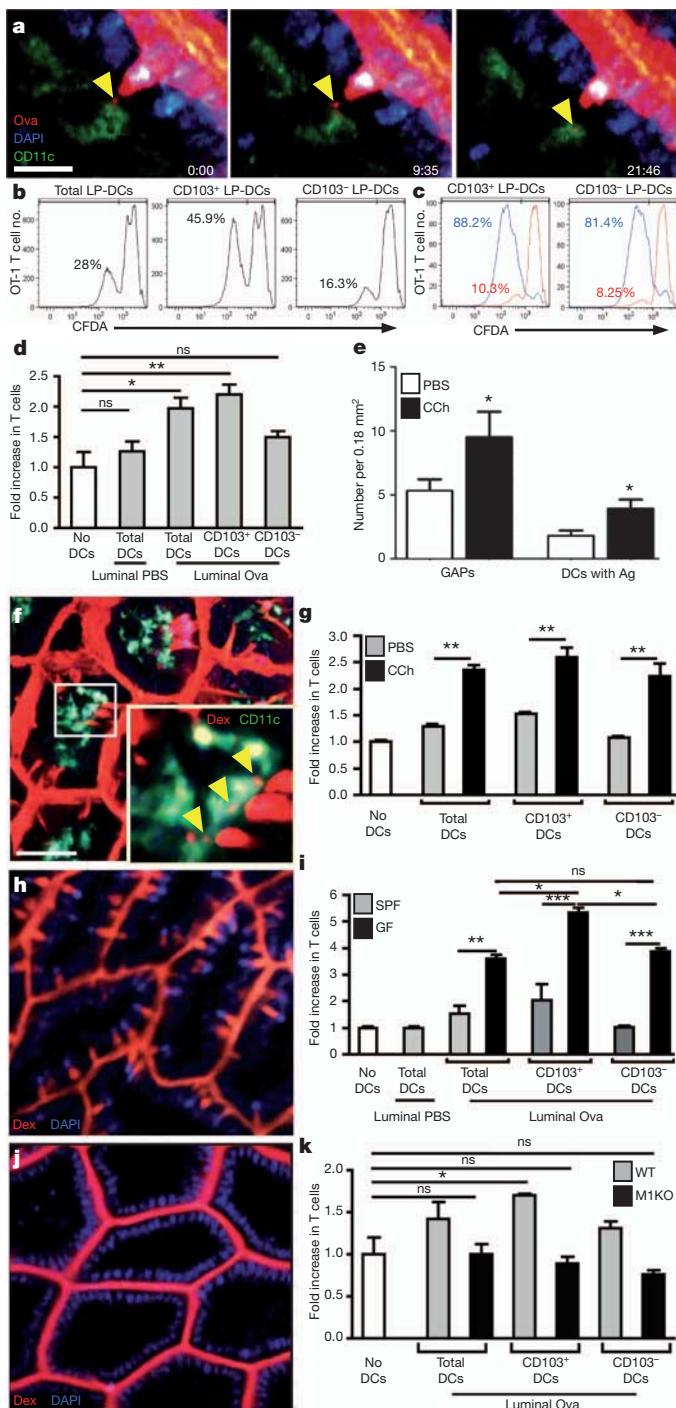


Figure 4 | GAPs are a source of luminal antigen for CD103⁺ LP-DCs in the steady-state. **a**, Two-photon time-lapse images of a LP-DC (CD11c-YFP⁺, green) acquiring fluorescent Ova (yellow arrow) from a GAP (red) in the intestinal epithelium (DAPI-stained nuclei, blue). Scale bar, 15 μ m; time stamp, min:s of elapsed time. **b–d**, LP-DC antigen presentation capacity in mice given luminal Ova or PBS assessed on day 3 by CFDA dilution (**b**, **c**) and by counting the number of T cells after co-culture with LP-DCs (**d**). **c**, CD103[–] LP-DCs induced T-cell proliferation comparable to CD103⁺ LP-DCs when exogenous antigen was added to the *in vitro* cultures (blue histogram), PBS controls (red histograms). **e**, **f**, Two-photon imaging of GAPs and LP-DCs in CCh-treated mice. CCh increased the number of GAPs and the colocalization of 10 kDa dextran (red) with LP-DCs (green). **f**, Inset, LP-DCs capturing dextran (yellow arrowheads). **g**, Luminal antigen presentation by LP-DCs following CCh administration as compared to controls given luminal Ova. **h**, Germ-free mice had increased GAPs (10 kDa dextran, red; DAPI, blue). **i**, Luminal antigen presentation by LP-DCs from germ-free mice was significantly enhanced compared to LP-DCs from SPF housed mice. CD103⁺ LP-DCs presented luminal antigen significantly better than CD103[–] LP-DCs from germ-free mice. **j**, Two-photon imaging of intestines in M1KO mice. M1KO mice lacked GAPs (red, 10 kDa dextran) in the epithelium (blue, DAPI). **k**, Luminal antigen presentation by LP-DCs from M1KO mice was undetectable. Scale bar, 20 μ m (**a**), 40 μ m (**f**). ns, not significant. * P < 0.05, ** P < 0.01, *** P < 0.001. Error bars, s.e.m. Data in **d**, **g** and **k** are representative of three or more independent experiments, data in **i** was representative of two independent experiments. In **e**, n = 15 or more images were obtained from three animals for each condition.

given luminal Ova showed significantly increased antigen presentation capacity compared to LP-DCs from SPF mice. Moreover, luminal antigen delivery in germ-free mice favoured CD103⁺ LP-DCs over CD103[–] LP-DCs (Fig. 4i), similar to observations in SPF mice.

To test directly whether GAPs are a major pathway of antigen delivery to LP-DCs, we generated mice with an epithelial-cell-specific deletion in mouse atonal homologue 1 (Math1), a transcription factor required for the development of secretory cell lineages, including GCs³². Math1^{fl/fl}Cre mice (M1KO) lacked small intestine GCs (Supplementary Fig. 9a–c) and *in vivo* two-photon imaging revealed a concomitant loss of GAPs, with the epithelium forming a tight barrier to luminal dextran (Fig. 4j and Supplementary Fig. 9d). Moreover, LP-DC populations isolated from M1KO mice given luminal Ova failed to stimulate OTI T cells above background levels (Fig. 4k and Supplementary Fig. 9e). The lack of antigen presentation by CD103⁺ LP-DCs from M1KO mice was not due to an intrinsic defect, because these LP-DCs were capable of inducing robust T-cell proliferation when exogenous Ova was added to the T-cell cultures (Supplementary Fig. 9f). Taken together, these findings indicate that GAPs are a major mechanism for delivering luminal antigens to LP-DCs in the steady state.

Understanding how the balance between tolerance and immunity is achieved at the intestinal mucosa is crucial for oral vaccine development and the treatment of chronic intestinal inflammatory diseases. This study identifies GAPs as a mechanism by which CD103⁺ LP-DCs can acquire innocuous antigens from the intestinal lumen in the steady state. How preferential antigen delivery is achieved is the focus of ongoing studies. Biased delivery could be a natural consequence of CD103⁺ LP-DC sampling the epithelium more actively or perhaps being recruited selectively by chemoattractants released by or near GCs. GC deficiency or dysfunction in mice and humans has been linked to the development of intestinal inflammation^{23–27}. Although this association has been attributed to the loss of mucins and other biologically active GC products, our findings suggest that GCs could play a key role in promoting intestinal immune homeostasis by delivering luminal antigen to tolerogenic LP-DCs.

METHODS SUMMARY

Mice and human specimens. All mice were of a C57BL/6 background unless otherwise specified. Healthy human jejunal sections were obtained from bariatric surgeries, placed in PBS containing dextran for 1 h before imaging. Procedures and protocols were carried out in accordance with the institutional review board at Washington University School of Medicine.

Two-photon microscopy. The intravital imaging preparation used in this study is similar to previously described methods^{11,15} with the following differences: imaging is performed with the tissue within the peritoneal cavity, faecal material is not scraped from the mucosal surface and in some experiments atropine (1 mg kg⁻¹) was injected subcutaneously to dampen peristaltic movement of the small intestine. At this dose atropine did not affect the formation of TEDs or the detection of GAPs. Model fluorescent antigens, dextran (2–5 mg), Ova (2 mg), BSA, (2 mg) and FluoSpheres (1 ml undiluted) (all from Invitrogen) were injected into the intestinal lumen approximately 2 h before imaging. Human resection specimens were incubated in 10 µg ml⁻¹ of dextran at room temperature for 1 h before imaging.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 August 2011; accepted 17 January 2012.

1. Weiner, H. L., da Cunha, A. P., Quintana, F. & Wu, H. Oral tolerance. *Immunol. Rev.* **241**, 241–259 (2011).
2. Jaenson, E. et al. Small intestinal CD103⁺ dendritic cells display unique functional properties that are conserved between mice and humans. *J. Exp. Med.* **205**, 2139–2149 (2008).
3. Johansson-Lindbom, B. Functional specialization of gut CD103⁺ dendritic cells in the regulation of tissue-selective T cell homing. *J. Exp. Med.* **202**, 1063–1073 (2005).
4. Uematsu, S. et al. Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. *Nature Immunol.* **9**, 769–776 (2008).
5. Varol, C. et al. Intestinal lamina propria dendritic cell subsets have different origin and functions. *Immunity* **31**, 502–512 (2009).
6. Schulz, O. et al. Intestinal CD103⁺, but not CX3CR1⁺, antigen sampling cells migrate in lymph and serve classical dendritic cell functions. *J. Exp. Med.* **206**, 3101–3114 (2009).
7. Bogunovic, M. et al. Origin of the lamina propria dendritic cell network. *Immunity* **31**, 513–525 (2009).
8. Coombes, J. L. et al. A functionally specialized population of mucosal CD103⁺ DCs induces Foxp3⁺ regulatory T cells via a TGF-β and retinoic acid-dependent mechanism. *J. Exp. Med.* **204**, 1757–1764 (2007).
9. Sun, C. M. et al. Small intestine lamina propria dendritic cells promote de novo generation of Foxp3 T reg cells via retinoic acid. *J. Exp. Med.* **204**, 1775–1785 (2007).
10. Niess, J. H. & Adler, G. Enteric flora expands gut lamina propria CX3CR1⁺ dendritic cells supporting inflammatory immune responses under normal and inflammatory conditions. *J. Immunol.* **184**, 2026–2037 (2010).
11. Watson, A. J. et al. Epithelial barrier function *in vivo* is sustained despite gaps in epithelial layers. *Gastroenterology* **129**, 902–912 (2005).
12. Hesse, M. et al. A mutation of keratin 18 within the coil 1A consensus motif causes widespread keratin aggregation but cell type-restricted lethality in mice. *Exp. Cell Res.* **313**, 3127–3140 (2007).
13. Hase, K. et al. Uptake through glycoprotein 2 of FimH⁺ bacteria by M cells initiates mucosal immune response. *Nature* **462**, 226–230 (2009).
14. Kühn, R., Lohler, J., Rennick, D., Rajewsky, K. & Müller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
15. Chieppa, M., Rescigno, M., Huang, A. Y. & Germain, R. N. Dynamic imaging of dendritic cell extension into the small bowel lumen in response to epithelial cell TLR engagement. *J. Exp. Med.* **203**, 2841–2852 (2006).
16. Niess, J. H. CX₃CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* **307**, 254–258 CrossRef (2005).
17. Rescigno, M. et al. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nature Immunol.* **2**, 361–367 (2001).
18. Rescigno, M., Rotta, G., Valzasina, B. & Ricciardi-Castagnoli, P. Dendritic cells shuttle microbes across gut epithelial monolayers. *Immunobiology* **204**, 572–581 (2001).
19. Shen, L., Weber, C. R., Raleigh, D. R., Yu, D. & Turner, J. R. Tight junction pore and leak pathways: a dynamic duo. *Annu. Rev. Physiol.* **73**, 283–309 (2011).
20. Madara, J. L. & Trier, J. S. Structure and permeability of goblet cell tight junctions in rat small intestine. *J. Membr. Biol.* **66**, 145–157 (1982).
21. Hogquist, K. A. et al. T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
22. Shroyer, N. F. et al. Intestine-specific ablation of *Mouse atonal homolog 1* (*Math1*) reveals a role in cellular homeostasis. *Gastroenterology* **132**, 2478–2488 (2007).
23. Dvorak, A. M. & Dickerin, G. R. Crohn's disease: transmission electron microscopic studies. I. Barrier function. Possible changes related to alterations of cell coat, mucous coat, epithelial cells, and Paneth cells. *Hum. Pathol.* **11**, 561–571 (1980).
24. Eri, R. D. et al. An intestinal epithelial defect conferring ER stress results in inflammation involving both innate and adaptive immunity. *Mucosal Immunol.* **4**, 354–364 (2011).
25. Trabucchi, E. et al. Differential diagnosis of Crohn's disease of the colon from ulcerative colitis: ultrastructure study with the scanning electron microscope. *Int. J. Tissue React.* **8**, 79–84 (1986).
26. Tytgat, K. M., van der Wal, J. W., Einerhand, A. W., Buller, H. A. & Dekker, J. Quantitative analysis of MUC2 synthesis in ulcerative colitis. *Biochem. Biophys. Res. Commun.* **224**, 397–405 (1996).
27. Van der Sluis, M. et al. Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117–129 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by grants DK064798 (R.D.N.), AI083538 (R.D.N.), AI095550 (R.D.N. and M.J.M.), DK085941 (L.W.W.) and AI077600 (M.J.M.). The authors thank W. Beatty for assistance with confocal microscopy, C. Eagon for assistance with human specimens, the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis for the use of the Siteman Flow Cytometry Core, which provided high-speed flow sorting and the Washington University Digestive Disease Research Core Center (DDRCC), which provided gnotobiotic mice. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant number P30 CA91842. The Washington University DDRCC is supported by grant P30-DK52574.

Author Contributions J.R.M. and L.W.W. contributed equally to this work. J.R.M., B.W. and V.K. performed two-photon imaging experiments and data analysis, L.W.W., K.A.K. and K.G.M. performed cell isolation, *in vitro* studies and immunofluorescence and data analysis, R.D.N. and M.J.M. conceived of the study, directed the experimental design, analysed the data and wrote the manuscript. R.D.N. and M.J.M. contributed equally to this work and are equal corresponding authors. All authors reviewed and discussed the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.J.M. (miller@pathology.wustl.edu) or R.D.N. (rnewberry@wustl.edu).

METHODS

Mice and human specimens. C57BL/6 mice, *Rag*^{-/-} mice, Balb/c mice, *IL-10*^{-/-} knockout mice on the C3H/HejBir.129 background¹⁴, OTI T cell receptor transgenic mice²¹, CX₃CR1-GFP knock-in mice²⁸, Math1^{fl/fl} mice²² and Villin Cre transgenic mice²⁹ were purchased from The Jackson Laboratory. CD11c-YFP transgenic mice³⁰ were a gift from M. Nussenzweig and LysM-GFP mice³¹ were a gift from K. Ley. Gnotobiotic mice were obtained from the Washington University Digestive Disease Research Core Center murine models core. Animals, other than gnotobiotic mice, were housed in a specific-pathogen-free facility and fed routine chow diet. Animals were 8 to 16 weeks of age at the time of analysis. Intestines from mice receiving 250 rads of gamma irradiation and killed 6 h later served as positive controls for apoptosis markers. All mice were on a C57BL/6 background unless otherwise specified. Healthy human jejunal sections were obtained from bariatric surgeries, placed in PBS containing dextran for 1 h before imaging. Procedures and protocols were carried out in accordance with the institutional review board at Washington University School of Medicine.

Intravital two-photon microscopy. The intravital imaging preparation used in this study is similar to previously described methods^{11,15} with the following differences: the intestine is not exteriorized during imaging, faecal material is not scraped from the mucosal surface and in some experiments atropine (1 mg kg⁻¹) was injected subcutaneously to dampen peristaltic movement of the small intestine. Mice were anaesthetized with isoflurane and a small vertical incision was made in the abdominal wall to expose the peritoneal cavity and contents. The intestine was secured to the bottom of a glass coverslip on the upper chamber plate using a thin ring of Vetbond tissue adhesive (3M). Because the coverslip sits directly over the incision in the mouse abdomen the tissue remains in the peritoneal cavity for imaging. No additional manipulations were performed to image from the serosal surface. The imaging chamber was maintained at 37 °C using a dual-channel heating system (Warner Instruments). To image from the luminal surface, a small longitudinal incision was made in the intestine taking care to avoid large blood vessels. Model fluorescent antigens, dextran (2–5 mg), Ova (2 mg), BSA (2 mg) and FluoSpheres (1 ml undiluted) (all from Invitrogen) were injected into the intestinal lumen ~2 h before imaging. Human resection specimens were incubated in 10 µg ml⁻¹ of dextran at room temperature for 1 h before imaging. Time-lapse imaging was performed with a custom-built two-photon microscope running ImageWarp acquisition software (A&B Software). For time-lapse imaging, we averaged 15 video-rate frames (0.5 s per slice) during the acquisition. Each plane represents an image of 220 × 240 µm in the x and y dimensions. 21 to 31 sequential z-steps (2.5 µm each) were acquired to form a z-stack. In our experiments epithelial integrity was assessed by dextran and DAPI staining, which in healthy tissue, demarcates the luminal surface of the epithelium (aside from GAPs) and shows an ordered arrangement of DAPI-stained nuclei, respectively. In some experiments mice were given 10⁸ *Salmonella typhimurium* strain χ3716 orally 24–72 h before imaging.

Flow cytometry, immunohistochemistry and confocal microscopy. Flow cytometry and the staining of intestine sections were performed as previously described³². Reagents used for flow cytometry include anti-CD11c, anti-CD45, anti-CD4, anti-MHCII, anti-CD3, anti-CD8a, anti-PDCA-1 (all from eBioscience), 7-aminoactinomycin D (7-AAD), anti-CD103 and anti-B220 (both from BD Biosciences). Data was acquired with a FACScan cytometer (BD Biosciences) retrofitted with additional lasers. Data acquisition was performed using CellQuest (BD Biosciences) and Rainbow (Cytek) or FlowJo software (Tree Star). Data analysis was performed on a Macintosh computer running FlowJo software.

To immobilize the lysine-fixable fluorescent dextrans, intestinal sections were treated with 2% paraformaldehyde immediately after two-photon imaging. Reagents used for immunohistochemistry include 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich), anti-CD11c, anti-CD103, anti-e cadherin (all from BD Biosciences), anti-cytokeratin 18 and anti-MUC2 (both from abcam), anti-GP2 (MBL International), anti-cleaved cytokeratin 18 (Enzo Life Sciences), and anti-cleaved caspase 3 (Cell Signaling Technology). The terminal deoxynucleotidyl transferase-mediated dUTP nick end labelling (TUNEL) assay (Trevigen) was performed as per the manufacturers recommendations. Confocal microscopy was performed using a Zeiss LSM510 Meta laser scanning confocal microscope (Carl Zeiss) equipped with a ×63, 1.4 numerical aperture Zeiss Plan Apochromat oil objective. Images were obtained using the Zeiss LSM510 software.

T-cell proliferation assays. Mice were anaesthetized and the small intestine injected intraluminally with 2 mg of Ova (Sigma-Aldrich) dissolved in phosphate buffered saline (PBS), or PBS alone (controls) as in the two-photon imaging experiments. In some experiments mice received 3 µg of carbamylcholine (Sigma-Aldrich) subcutaneously 20 min after the administration of luminal Ova. Two hours after the administration of Ova, cell populations were isolated from the intestinal lamina propria as previously described³² and sorted by flow cytometry into total LP-DC populations (7-AAD⁻, CD45⁺ MHCII⁺ CD11c⁺), CD103⁺ DC populations (7-AAD⁻, CD45⁺ MHCII⁺ CD11c⁺ CD103⁺), or CD103⁻ DC populations (7-AAD⁻, CD45⁺ MHCII⁺ CD11c⁺ CD103⁻). Sorted DC populations were cultured with sorted CFDA (Invitrogen)-labelled OTI splenic T cells at a ratio of 1:10 DC to T cells. As a positive control, 20 µg of Ova was added to cultures of DC populations isolated from mice receiving luminal PBS, unless otherwise stated. After 3 days cultures were evaluated for CFDA dilution and the number of T cells by flow cytometry and cell counting.

Real-time PCR. RNA isolation, cDNA synthesis, standard curve construction and real time PCR were performed as previously described³². The following primers were used: 18S, forward, 5'-CGGCTACCACATCCAAGGAA-3' and reverse, 5'-GCTGGAATTACCGCGGCT-3'; cytokeratin 18, forward, 5'-CAGCCAGCG TCTATGCAGG-3' and reverse, 5'-CTTTCTCGGTCTGGATTCCAC-3'. Small intestine epithelial cell populations were stained with cytokeratin 18 and AlexaFluor 488 (Invitrogen)-labelled lectin from *Ulex europaeus* (UEA-I; Sigma-Aldrich). Goblet cell populations were isolated by flow cytometric sorting as UEA-I⁺ cytokeratin 18⁺ cells.

Statistical analysis. Data analysis using Student's *t* test or a one-way ANOVA with a Tukey's post test were performed using GraphPad Prism (GraphPad Software).

28. Jung, S. et al. Analysis of fractalkine receptor CX(3)CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell. Biol.* **20**, 4106–4114 (2000).
29. Madison, B. B. et al. *cis* elements of the villin gene control expression in restricted domains of the vertical (crypt) and horizontal (duodenum, cecum) axes of the intestine. *J. Biol. Chem.* **277**, 33275–33283 (2002).
30. Lindquist, R. L. et al. Visualizing dendritic cell networks *in vivo*. *Nature Immunol.* **5**, 1243–1250 (2004).
31. Faust, N., Varas, F., Kelly, L. M., Heck, S. & Graf, T. Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* **96**, 719–726 (2000).
32. McDonald, K. G., McDonough, J. S., Dieckgraefe, B. K. & Newberry, R. D. Dendritic cells produce CXCL13 and participate in the development of murine small intestine lymphoid tissues. *Am. J. Pathol.* **176**, 2367–2377 (2010).

Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human

Atsuhiko Ichimura^{1*}, Akira Hirasawa^{1*}, Odile Poulain-Godefroy^{2,3*}, Amélie Bonnefond^{2,3*}, Takafumi Hara¹, Loïc Yengo^{2,3}, Ikuo Kimura¹, Audrey Leloire^{2,3}, Ning Liu¹, Keiko Iida¹, Hélène Choquet^{2,3}, Philippe Besnard⁴, Cécile Lecoer^{2,3}, Sidonie Vivequin^{2,3}, Kumiko Ayukawa¹, Masato Takeuchi¹, Kentaro Ozawa¹, Maithé Tauber⁵, Claudio Maffei^{6,7}, Anita Morandi^{2,3,6}, Raffaella Buzzetti⁸, Paul Elliott⁹, Annely Pouta^{10,11}, Marjo-Riitta Jarvelin^{9,10,12}, Antje Körner¹³, Wieland Kiess¹³, Marie Pigeyre^{14,15}, Roberto Caiazzo^{14,16}, Wim Van Hul¹⁷, Luc Van Gaal¹⁸, Fritz Horber¹⁹, Beverley Balkau^{20,21}, Claire Lévy-Marchal²², Konstantinos Rouskas^{2,3,23}, Anastasia Kouvatsi²³, Johannes Hebebrand²⁴, Anke Hinney²⁴, Andre Scherag²⁵, François Pattou^{14,16}, David Meyre^{2,3,26}, Taka-aki Koshimizu²⁷, Isabelle Wolowczuk^{2,3}, Gozoh Tsujimoto¹ & Philippe Froguel^{2,3,28}

Free fatty acids provide an important energy source as nutrients, and act as signalling molecules in various cellular processes^{1–4}. Several G-protein-coupled receptors have been identified as free-fatty-acid receptors important in physiology as well as in several diseases^{3,5–13}. GPR120 (also known as O3FAR1) functions as a receptor for unsaturated long-chain free fatty acids and has a critical role in various physiological homeostasis mechanisms such as adipogenesis, regulation of appetite and food preference^{5,6,14–16}. Here we show that GPR120-deficient mice fed a high-fat diet develop obesity, glucose intolerance and fatty liver with decreased adipocyte differentiation and lipogenesis and enhanced hepatic lipogenesis. Insulin resistance in such mice is associated with reduced insulin signalling and enhanced inflammation in adipose tissue. In human, we show that GPR120 expression in adipose tissue is significantly higher in obese individuals than in lean controls. GPR120 exon sequencing in obese subjects reveals a deleterious non-synonymous mutation (p.R270H) that inhibits GPR120 signalling activity. Furthermore, the p.R270H variant increases the risk of obesity in European populations. Overall, this study demonstrates that the lipid sensor GPR120 has a key role in sensing dietary fat and, therefore, in the control of energy balance in both humans and rodents.

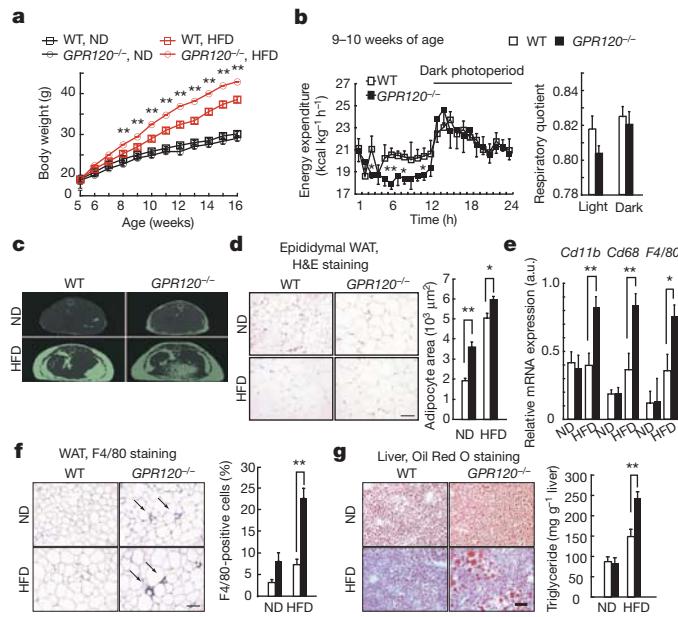
To investigate the role of GPR120 in metabolism, we examined GPR120-deficient mice (Supplementary Fig. 1) with respect to lipogenesis, glucose and energy homeostasis. On a normal diet containing 13% fat, the body weight was similar in both GPR120-deficient and wild-type mice. However, when 5-week-old GPR120-deficient mice were fed a high-fat diet (HFD) containing 60% fat, their body weight increase was ~10% higher than that of wild-type mice on a HFD (Fig. 1a). The difference in HFD-induced body weight gain between wild-type and GPR120-deficient mice was marked at ~8–10 weeks old and reached a plateau at 13 weeks old. To assess energy expenditure and substrate utilization, we next performed indirect calorimetry on wild-type and

mutant mice on a HFD at 9–10 weeks old (Fig. 1b) and 15–16 weeks old (Supplementary Fig. 2a). The young GPR120-deficient mice showed decreased energy expenditure compared with the young wild-type mice, particularly during the light/inactive phase (Fig. 1b, left), whereas older mutant and wild-type mice showed no such a difference (Supplementary Fig. 2a, left). The difference in energy expenditure between GPR120-deficient and wild-type mice was observed only in the light phase, indicating that the lack of the GPR120 receptor primarily affects basal metabolism, especially in young mice. The decreased energy expenditure might explain the difference we found in body weight gain between HFD-fed wild-type and mutant young mice. The lower values of respiratory quotient in mutant mice could be due to insufficient glucose utility, probably as a result of the decreased insulin sensitivity. In all experiments, both groups of mice showed similar levels of locomotor activity (data not shown).

White adipose tissue (WAT) and liver were substantially heavier in HFD-fed GPR120-deficient mice (Supplementary Fig. 2b). Plasma low- and high-density lipoprotein cholesterol levels were significantly higher in HFD-fed GPR120-deficient mice, and serum alanine aminotransferase levels were substantially increased, indicating abnormal cholesterol metabolism and liver function (Supplementary Table 1). Microcomputed tomography scanning revealed that 16-week-old GPR120-deficient mice stored much more fat than did wild type (Fig. 1c). A significant increase in adipocyte size in both epididymal (Fig. 1d) and subcutaneous (Supplementary Fig. 2c) fat was observed in GPR120-deficient mice. Furthermore, the expression of macrophage marker genes (*Cd11b* (*Itgam*), *Cd68* and *F4/80* (*Emr1*)) and the number of F4/80-positive cells were markedly enhanced in epididymal tissue from HFD-fed GPR120-deficient mice (Fig. 1e, f). Moreover, these mice showed liver steatosis and hepatic triglyceride content was markedly increased (Fig. 1g). Overall, HFD-induced obesity and liver fattiness were more severe in GPR120-deficient mice than in wild type.

¹Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan. ²Centre National de la Recherche Scientifique (CNRS)-Unité mixte de recherche (UMR) 8199, Lille Pasteur Institute, Lille 59000, France. ³Lille Nord de France University, Lille 59000, France. ⁴Institut National de la Santé et de la Recherche Médicale (Inserm)-UMR U866, Physiologie de la Nutrition, Bourgogne University, AgroSup Dijon, Dijon 21078, France. ⁵Inserm-U563, Children's Hospital, Centre Hospitalier Universitaire, Toulouse 31000, France. ⁶Regional Centre for Juvenile Diabetes, Obesity and Clinical Nutrition, Verona 37134, Italy. ⁷Department of Mother and Child, Biology-Genetics, Section of Paediatrics, University of Verona, Verona 37134, Italy. ⁸Department of Clinical Sciences, La Sapienza University, Rome 00161, Italy. ⁹Medical Research Council-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, St Mary's campus, Imperial College London, London W2 1PG, UK. ¹⁰National Public Health Institute, Biocenter Oulu, University of Oulu, Oulu 90220, Finland. ¹¹Institute of Clinical Medicine/Obstetrics and Gynecology, University of Oulu, Oulu 90220, Finland. ¹²Institute of Health Sciences, University of Oulu, Oulu 90220, Finland. ¹³Center for Pediatric Research, Department of Women's & Child Health, University of Leipzig, Leipzig 04317, Germany. ¹⁴Inserm-U859, Lille Nord de France University, Lille 59000, France. ¹⁵Lille University Hospital, Nutrition, Lille 59000, France. ¹⁶Lille University Hospital, Endocrine Surgery, Lille 59000, France. ¹⁷Department of Medical Genetics, University of Antwerp, Antwerp 2610, Belgium. ¹⁸Department of Endocrinology, Antwerp University Hospital, Antwerp 2650, Belgium. ¹⁹Department of Surgery and Internal Medicine, Clinic Lindberg, Medical Department, Winterthur 8400, and University of Berne, Berne 3011, Switzerland. ²⁰Inserm-U780, Centre for research in Epidemiology and Population Health (CRESP), Villejuif 94800, France. ²¹Paris-Sud 11 University, Orsay 91405, France. ²²Inserm-U690, Robert Debré hospital, Paris 75935, France. ²³Department of Genetics, Development and Molecular Biology, School of Biology, Aristotle University of Thessaloniki, Thessaloniki 541 24, Greece. ²⁴Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Essen 45147, Germany. ²⁵Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen 45122, Germany. ²⁶McMaster University, Hamilton, Ontario L8S 4L8, Canada. ²⁷Department of Pharmacology, Division of Molecular Pharmacology, Jichi Medical University, Tochigi 329-0498, Japan. ²⁸Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London W12 0NN, UK.

*These authors contributed equally to this work.



Obesity-associated insulin resistance was also more severe in GPR120-deficient mice. HFD-fed GPR120-deficient mice showed higher levels of fasting plasma glucose and insulin than did wild type, although these parameters were similar between the two groups on a normal diet (Fig. 2a). HFD-induced insulin resistance, as determined by an insulin tolerance test, was more prominent in GPR120-deficient mice than in wild type (Fig. 2b, left, and Supplementary Fig. 3a, b). A glucose tolerance test further revealed that these mice suffered from impaired glucose metabolism (Fig. 2b, right, and Supplementary

Fig. 3a, b). The level of plasma leptin was significantly higher in HFD-fed GPR120-deficient mice than in wild type (Supplementary Fig. 3c). However, there was no significant difference in terms of plasma adiponectin level or food intake between the two groups (Supplementary Fig. 3d, e). HFD-fed GPR120-deficient mice showed a marked increase in the size of islets and KI67 (MKI67)-positive cells, suggesting adaptive enlargement of the β -cell mass in response to insulin resistance^{17,18} (Supplementary Fig. 3f, g). Moreover, we observed markedly reduced peripheral insulin sensitivity in tissues from HFD-fed GPR120-deficient mice (Fig. 2c). Insulin was shown to induce the phosphorylation of AKT (AKT1) (on Ser 473) in WAT, liver and skeletal muscle, with similar intensities in wild-type and GPR120-deficient mice on a normal diet (Supplementary Fig. 3h). Consistent with the insulin resistance reported above, HFD-fed GPR120-deficient mice showed loss of insulin-induced AKT phosphorylation in WAT and the liver.

To determine the molecular basis of the metabolic changes in WAT and livers of GPR120-deficient mice, we performed gene expression

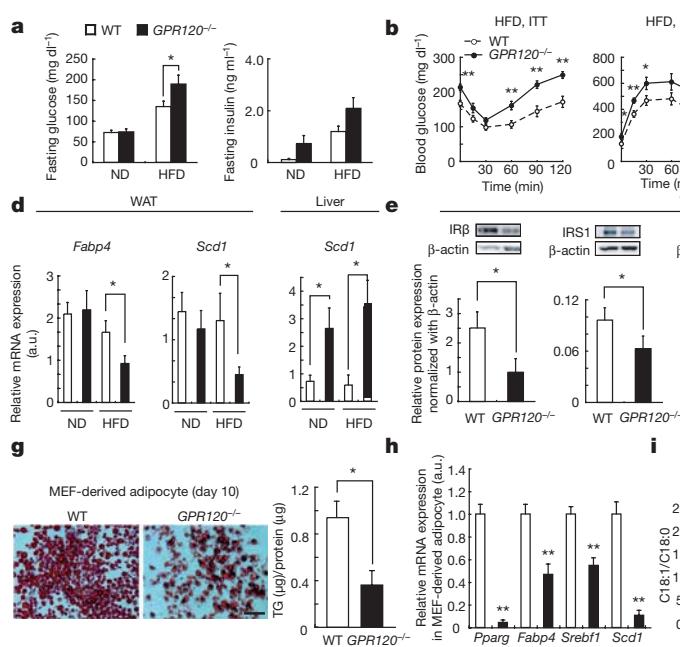
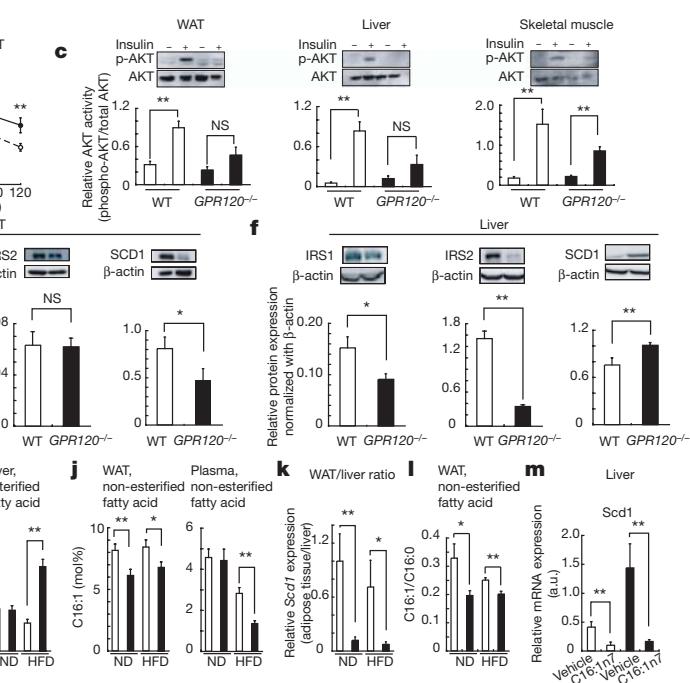


Figure 2 | Impaired glucose metabolism, adipogenesis and lipogenesis in HFD-fed GPR120-deficient mice. a, Fasting blood glucose and serum insulin levels ($n = 6–15$). b, Plasma glucose during insulin tolerance test (ITT, left) and glucose tolerance test (GTT, right) ($n = 12–14$). c, Phosphorylation of AKT (Ser 473) in WAT, liver and skeletal muscle after 24 h fasting ($n = 6, 7$). NS, not significant. d, Relative mRNA expression of *Fabp4* and *Scd1* in WAT or *Scd1* in liver ($n = 6$). e, Protein expression of $\text{IR}\beta$, IRS1, IRS2, SCD1 and β -actin in WAT. f, Protein expression of IRS1, IRS2, SCD1 and β -actin in liver. g, Oil Red O-staining and triglyceride (TG) content of mouse embryonic



fibroblast (MEF)-derived adipocyte. Scale bar, 50 μm . h, Relative mRNA expression in MEF-derived adipocyte ($n = 6$). i, The ratio of C18:1 to C18:0 in livers ($n = 6–8$). j, Non-esterified C16:1n7 palmitoleate in WAT and plasma ($n = 4–7$). k, The ratio of *Scd1* mRNA expression in liver and WAT ($n = 6, 7$). l, The ratio of C16:1 to C16:0 in adipose tissues ($n = 6–8$). m, Hepatic *Scd1* mRNA expression in mice infused with vehicle or triglyceride:palmitoleate for 6 h ($n = 4, 5$). All data represent mean \pm s.e.m. * $P < 0.05$ and ** $P < 0.01$ versus the corresponding wild-type value.

analyses. We identified approximately 700 differentially expressed genes in WAT between HFD-fed GPR120-deficient and wild-type mice (Supplementary Fig. 4a). Connectivity mapping of these genes showed that pathways relating to insulin signalling and adipocyte differentiation were depressed, whereas those related to inflammation were enhanced in HFD-fed GPR120-deficient mice (Supplementary Fig. 4b). Quantitative real-time PCR (qRT-PCR) analysis confirmed the downregulation of insulin-signalling-related genes (*Insr*, *Irs1* and *Irs2*), an adipocyte differentiation marker gene (*Fabp4*) and a lipogenesis-related gene (*Scd1*) in the epididymal fat from HFD-fed GPR120-deficient mice (Fig. 2d and Supplementary Fig. 3i). We identified approximately 100 differentially expressed genes in the liver between HFD-fed GPR120-deficient and wild-type mice (Supplementary Fig. 5). Notably, lipogenesis-related genes (*Scd1* and *Me1*) and a fatty acid transporter gene (*Cd36*) were significantly upregulated in livers from GPR120-deficient mice. Quantitative RT-PCR analysis confirmed upregulation of *Scd1* in the liver of GPR120-deficient mice (Fig. 2d).

Western blot analysis further confirmed downregulation of IR β , IRS1 and SCD1 in adipose tissue of HFD-fed GPR120-deficient mice (Fig. 2e) but downregulation of IRS1 and IRS2 and upregulation of SCD1 in their livers (Fig. 2f). Hence, insulin-signalling-related molecules were downregulated by the lack of GPR120 in both adipose tissue and the liver. However, the expression of SCD1, the rate-limiting enzyme in the biosynthesis of mono-unsaturated fatty acids, was downregulated in adipose tissue but upregulated in liver. Furthermore, the expression of *Scd1* and several adipogenic genes^{14,19} (*Pparg*, *Fabp4* and *Srebf1*) was suppressed in mouse-embryonic-fibroblast-derived adipocytes from GPR120-deficient mice, indicating that GPR120 is required for normal adipogenesis, as previously reported in differentiating 3T3-L1 adipocytes depleted of endogenous GPR120 by short interfering RNA¹⁴ (Fig. 2g, h).

To determine the effects of altered lipogenesis on lipid composition in GPR120-deficient mice, we performed lipidomics analysis in WAT, livers and plasma. Significant changes of major lipid clusters were observed (Supplementary Fig. 6). Notably, the hepatic concentration of oleate (C18:1n9c) was significantly higher in HFD-fed GPR120-deficient mice than in wild type. The ratio of C18:1 to C18:0, an indicator of SCD1 enzyme activity^{20–23}, was markedly enhanced in livers from HFD-fed GPR120-deficient mice relative to wild type (Fig. 2i). Moreover, the levels of C16:1n7 palmitoleate, which has recently been recognized as a lipid hormone⁴, in WAT and plasma were significantly lower in HFD-fed GPR120-deficient mice than in wild type (Fig. 2j). In particular, lower levels of C16:1n7 palmitoleate were detected even in WAT of GPR120-deficient mice on a normal diet (Fig. 2j), which seems to be in good agreement with the suppressed *Scd1* expression and the reduced SCD1 desaturation index^{20–23} (C16:1/C16:0; Fig. 2k, l). Lipidomics analysis clearly illustrated dysregulated lipogenesis in GPR120-deficient mice, and showed the reduced production of lipid hormone C16:1n7 palmitoleate⁴. To determine whether the enhanced hepatic lipogenesis in GPR120-deficient mice was due to the reduced levels of C16:1n7 palmitoleate, we examined the effect of C16:1n7 palmitoleate treatment on hepatic *Scd1* expression. A 6-h infusion of triglyceride:palmitoleate markedly lowered the enhanced hepatic *Scd1* expression in GPR120-deficient mice (Fig. 2m). The results indicated that the reduced C16:1n7 palmitoleate may explain the systemic metabolic disorders observed in GPR120-deficient mice on a HFD, as palmitoleate has been proposed to be a bioactive lipid by which adipose tissue communicates with distant organs (such as liver) and regulates systemic metabolic homeostasis⁴. This study shows that dysfunction of GPR120 can be an underlying mechanism for diet-associated obesity and obesity-related metabolic disorders in mouse.

The mouse data prompted us to assess the potential contribution of GPR120 to the development of obesity and its metabolic complications in humans. First, the expression levels of *GPR120* in both subcutaneous and omental adipose tissues as well as in liver samples were compared between lean and obese subjects. Normoglycaemic obese patients and

lean individuals ($n = 14$ in each group) were matched for age and gender (Supplementary Table 2). As previously described^{5,14}, we confirmed that *GPR120* is barely expressed in liver of either lean or obese subjects (data not shown). By contrast, we found that *GPR120* is well expressed in the adipose tissue of lean individuals (Fig. 3a). In addition, human obesity is significantly associated with an increase in *GPR120* expression in both subcutaneous and omental adipose tissues (1.8-fold increase; $P = 0.0004$ and $P = 0.003$, respectively). We also found that *GPR120* expression in subcutaneous adipose tissue strongly correlates with that in omental adipose tissue (Spearman analysis; $r = 0.570$ and $P = 2.74 \times 10^{-8}$), suggesting a systemic regulation of its expression in humans. Furthermore, we found a positive correlation between *GPR120* expression in both subcutaneous and omental adipose tissues and in subjects' concentrations of plasma low-density lipoproteins (on adjustment for age and sex; $r = 0.247$, $P = 0.0115$ and $r = 0.255$, $P = 0.0118$, respectively).

To investigate the contribution of *GPR120* to human obesity, the four *GPR120* exons were sequenced in 312 French, non-consanguineous, extremely obese children and adults (Supplementary Table 3). We

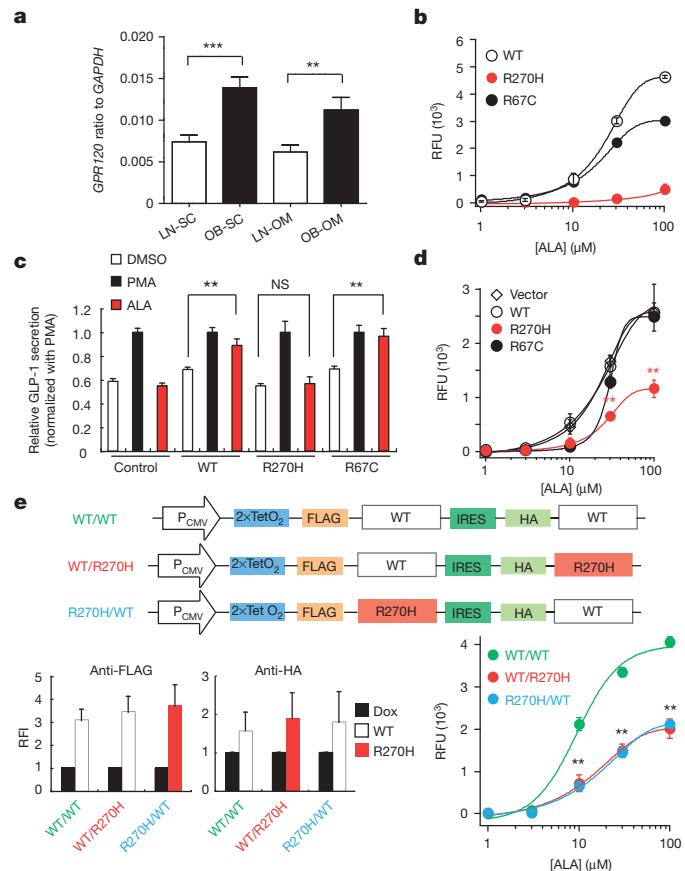


Figure 3 | GPR120 expression in human obese tissue samples, and effect of GPR120 variants on $[Ca^{2+}]_i$ response and GLP-1 secretion. **a**, *GPR120* mRNA levels in human subcutaneous (SC) and omental (OM) adipose tissues of lean (LN; $n = 14$) and obese (OB; $n = 14$) normoglycaemic individuals. Mann-Whitney analysis, *** $P = 0.0004$ and ** $P = 0.003$. **b**, ALA-induced $[Ca^{2+}]_i$ responses in cells expressing wild-type GPR120 or a p.R67C or p.R270H variant. **c**, ALA-induced GLP-1 secretion in NCI-H716 cells expressing a wild-type GPR120, a p.R67C or a p.R270H receptor. **d**, Effect of transfection with GPR120 variants on ALA-induced $[Ca^{2+}]_i$ response in cells stably expressing wild-type GPR120. **e**, Effect of co-expression of human GPR120 p.R270H variant with wild-type GPR120 on ALA-induced $[Ca^{2+}]_i$ response. Top: schematic diagram of constructs. Bottom: expression of wild type and p.R270H (left), and concentration- $[Ca^{2+}]_i$ response for ALA in cells expressing wild-type/wild-type, wild-type/R270H or R270H/wild-type receptors (right). ** $P < 0.01$ versus the corresponding control value. RFI, relative fluorescence intensity; RFU, relative fluorescence unit. All data show mean \pm s.e.m.

Table 1 | Identified variants in *GPR120* exons and association between the p.R67C/rs6186610 or p.R270H non-synonymous variant and obesity

	Variant	Nucleotide change	Chr 10 position	MAF ₁	MAF ₂ (controls, n = 7,654)	MAF ₂ (cases, n = 6,942)	Adjustment: age and gender		Adjustment: age, gender and geography	
							OR [95% CI]	P	OR [95% CI]	P
Missense variants	p.R67C/rs6186610	C → T	95,316,666	0.05	0.043	0.055	1.16 [1.02, 1.31]	0.022	1.13 [1.00, 1.28]	0.060
	p.R270H	G → A	95,337,031	0.03	0.013	0.024	1.62 [1.31, 2.00]	8.00 × 10 ⁻⁶	1.58 [1.28, 1.95]	2.17 × 10 ⁻⁵
Synonymous variants	p.V38V	G → A	95,316,581	0.0016	—	—	—	—	—	—
	p.S192S	G → A	95,325,846	0.0016	—	—	—	—	—	—
	p.V243V	C → T	95,328,938	0.0016	—	—	—	—	—	—
	p.S264S	G → A	95,337,014	0.0016	—	—	—	—	—	—

Variant position was indicated according to human genome build NCBI36/hg18. Association between p.R67C/rs6186610 or p.R270H variant and obesity was assessed by using a logistic regression adjusted for age and gender or for age, gender and geography, under an additive model. Chr, chromosome; MAF₁, minor allele frequency in sequencing data set (n = 312 extremely obese individuals); MAF₂, minor allele frequency in the large obesity case-control genotyping data set; OR, odds ratio; CI, confidence interval.

identified only two non-synonymous variants, R270H (minor allele frequency (MAF), ~3%) and p.R67C/rs6186610 (MAF, ~5%), and four rare synonymous variants (MAF, <1%) (Table 1). The two non-synonymous variants were subsequently genotyped in 6,942 unrelated obese individuals and 7,654 control subjects, all of European origin (Supplementary Table 4). By using a logistic regression model adjusted for age and sex, we found that R270H associated with obesity under an additive model (OR = 1.62 [1.31, 2.00]_{95%} (odds ratio and 95% confidence interval), P = 8.00 × 10⁻⁶; Table 1); whereas we found only a trend for association between p.R67C and obesity (OR = 1.16 [1.02, 1.31]_{95%}, P = 0.022; Table 1). It is noteworthy that these results were almost the same after adjusting for geographical origin (Table 1).

We then genotyped the p.R270H variant in 1,109 French pedigrees selected for obesity (n = 5,045) and in 780 German trios with one obese child (n = 2,340). We observed a significant over-transmission of the p.R270H low-frequency variant to obese offspring in 117 pedigrees or trios where the p.R270H variant was present (transmission, 62%; P = 0.009; Supplementary Table 5). This family-based study excludes a hidden population stratification effect as a cause of spurious association.

We assessed the functional significance of both the p.R67C mutation and the p.R270H mutation *in silico* using several programs: arginine residues at positions 67 and 270 presented a high-evolutionary-conservation pattern among mammals and the two amino-acid substitutions were predicted to be potentially damaging (Supplementary Table 6). To examine the influences of the two non-synonymous variants on GPR120 function *in vitro*, we assessed each receptor ability to mobilize intracellular calcium (concentration, [Ca²⁺]_i) in response to the endogenous agonist α-linolenic acid (ALA). We found that ALA induced [Ca²⁺]_i responses in T-REx 293 cells expressing either wild-type or p.R67C receptor in a dose-dependent manner, whereas ALA-induced [Ca²⁺]_i responses in cells expressing p.R270H were significantly lower (P = 1.6 × 10⁻⁵) than those in cells expressing wild type at ALA concentrations greater than 10 μM (Fig. 3b). We further examined the functional ability of the mutated receptors to secrete GLP-1 (ZGLP1) from human intestinal NCI-H716 cells, as this cell line lacks GPR120 expression and it can secrete GLP-1 in a regulated manner⁵. ALA induced secretion of GLP-1 in NCI-H716 cells expressing either wild-type (P = 0.004) or p.R67C (P = 3.2 × 10⁻⁵) receptor, but not in NCI-H716 cells expressing p.R270H mutant (P = 0.96) (Fig. 3c). The transfection efficiencies for the GPR120 variant receptors were confirmed to be almost the same (data not shown). To examine the effect of the p.R270H variant on the wild-type receptor signalling, we analysed the [Ca²⁺]_i dose-response curves after the transfection of an empty vector, a wild-type receptor plasmid or a p.R270H-mutated plasmid into T-REx 293 cells expressing wild-type GPR120. The transfection of the p.R270H-mutated plasmid suppressed dose-response curves, and the maximal ALA-induced [Ca²⁺]_i response was significantly decreased (P = 0.004; Fig. 3d).

To assess the effect more quantitatively, we analysed [Ca²⁺]_i dose-response curves in T-REx 293 cells stably expressing bicistronic wild-type/wild-type, wild-type/p.R270H or p.R270H/wild-type receptors

(Fig. 3e, top). Almost equal levels of receptor protein expression in each cell line were confirmed by flow cytometry analysis (Fig. 3e, bottom left). Compared with cells expressing wild-type/wild-type receptor, the [Ca²⁺]_i dose-response curves obtained in cells expressing either wild-type/p.R270H or p.R270H/wild-type receptor were markedly suppressed, and the maximal ALA-induced [Ca²⁺]_i response was significantly decreased (P = 1.2 × 10⁻⁵; Fig. 3e, bottom right). These findings suggest that the p.R270H variant that is significantly associated with obesity has an inhibitory effect on GPR120. The p.R270H mutant lacks the ability to transduce the signal of long-chain free fatty acids, contrary to the p.R67C mutant, which did not associate with obesity.

To analyse whether being a p.R270H variant carrier may affect GPR120 expression in the adipose tissue, we quantified GPR120 expression in samples from both obese p.R270H carriers and obese non-carriers. Two hundred and thirty-eight obese normoglycaemic patients from the Atlas Biologique de l'Obésité Sévère cohort had already been genotyped for the p.R270H variant. Ten subjects heterozygous for the p.R270H variant were matched for age, gender and body mass index with ten non-carrier (wild-type) obese normoglycaemic patients (Supplementary Table 7). The expression of GPR120 was similar between p.R270H carriers and wild-type subjects, both in subcutaneous and omental adipose tissues (Supplementary Fig. 7a), suggesting that the presence of the functionally deleterious mutation has no primary or secondary effect on gene expression in fat depots. The adipogenesis marker PPARG, the lipogenesis-related factor SCD and the macrophage marker CD68 were found similarly well expressed in the adipose tissues of wild-type and p.R270H carrier patients (Supplementary Fig. 7b, c). Nevertheless, the expression of the fatty-acid-binding protein FABP4 in omental adipose tissue was significantly lower in p.R270H carriers than in wild-type individuals (28% decrease, P = 0.043; Supplementary Fig. 7b).

Our results provide convincing evidence that the lipid sensor GPR120 is involved in obesity in both mice and humans. Given the role of GPR120 as a physiologic integrator of the environment (especially the fatty diet), these data provide insight into the molecular mechanisms by which the 'Westernized' diet may contribute to early-onset obesity and associated complications including non-alcoholic steatohepatitis. It also brings some understanding of the metabolic effects of the omega-3 fatty acids, which are often proposed as food supplements. This may open novel avenues of research for drug development in the treatment of obesity, lipid metabolism abnormalities and liver diseases, because receptors that sense free fatty acids represent attractive drug targets.

METHODS SUMMARY

GPR120-deficient mice were generated by deleting *Gpr120* exon 1. All animal procedures and euthanasia were reviewed by the local animal care committee approved by local government authorities. Blood analysis, extraction and detection of mRNA and proteins, and immunohistochemical analysis, were performed following standard protocols as described previously^{5,24–26}. Details of antibodies, primers and probes are given in Methods. The level of significance for the difference between data sets was assessed using Student's *t*-test. Analysis of variance followed by Tukey's test was used for multiple comparisons.

In human, *GPR120* expression in liver and in both omental and subcutaneous adipose tissues was assessed by quantitative RT-PCR (Taqman), in lean and obese subjects from the Atlas Biologique de l'Obésité Sévère cohort. The four *GPR120* exons were sequenced in 312 French, extremely obese subjects following a standard Sanger protocol. The two identified non-synonymous variants (p.R270H and p.R67C/rs186610) were subsequently genotyped in a large European obesity case-control study ($n_{\text{cases}} = 6,942$, $n_{\text{controls}} = 7,654$), by high-resolution melting analysis and TaqMan, respectively. The association between obesity status and each variant was assessed using logistic regression adjusted first for age and gender and then for age, gender and geography origin, under an additive model. The consequences of the two identified non-synonymous variants for *GPR120* function ($[\text{Ca}^{2+}]_i$ response and GLP-1 secretion) were assessed *in vitro*. The human study protocol was approved by the local ethics committee, and participants from all of the studies signed an informed consent form.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 November; accepted 14 December 2011.

Published online 19 February 2012.

1. Nunez, E. A. Biological complexity is under the 'strange attraction' of non-esterified fatty acids. *Prostaglandins Leukot. Essent. Fatty Acids* **57**, 107–110 (1997).
2. Haber, E. P. *et al.* Pleiotropic effects of fatty acids on pancreatic β -cells. *J. Cell. Physiol.* **194**, 1–12 (2003).
3. Itoh, Y. *et al.* Free fatty acids regulate insulin secretion from pancreatic beta cells through *GPR40*. *Nature* **422**, 173–176 (2003).
4. Cao, H. *et al.* Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell* **134**, 933–944 (2008).
5. Hirasawa, A. *et al.* Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through *GPR120*. *Nature Med.* **11**, 90–94 (2005).
6. Steneberg, P., Rubins, N., Bartoo, Shifman, R., Walker, M. D. & Edlund, H. The FFA receptor *GPR40* links hyperinsulinemia, hepatic steatosis, and impaired glucose homeostasis in mouse. *Cell Metab.* **1**, 245–258 (2005).
7. Wang, J., Wu, X., Simonavicius, N., Tian, H. & Ling, L. Medium-chain fatty acids as ligands for orphan G protein-coupled receptor *GPR84*. *J. Biol. Chem.* **281**, 34457–34464 (2006).
8. Ichimura, A., Hirasawa, A., Hara, T. & Tsujimoto, G. Free fatty acid receptors act as nutrient sensors to regulate energy homeostasis. *Prostaglandins Other Lipid Mediat.* **89**, 82–88 (2009).
9. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor *GPR43*. *Nature* **461**, 1282–1286 (2009).
10. Ahmed, K. *et al.* An autocrine lactate loop mediates insulin-dependent inhibition of lipolysis through *GPR81*. *Cell Metab.* **11**, 311–319 (2010).
11. Oh, Da, Y. *et al.* *GPR120* is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell* **142**, 687–698 (2010).
12. Hara, T., Hirasawa, A., Ichimura, A., Kimura, I. & Tsujimoto, G. Free fatty acid receptors *FFAR1* and *GPR120* as novel therapeutic targets for metabolic disorders. *J. Pharm. Sci.* **100**, 3594–3601 (2011).
13. Kimura, I. *et al.* Short-chain fatty acids and ketones directly regulate sympathetic nervous system via G protein-coupled receptor 41 (*GPR41*). *Proc. Natl. Acad. Sci. USA* **108**, 8030–8035 (2011).
14. Gotoh, C. *et al.* The regulation of adipogenesis through *GPR120*. *Biochem. Biophys. Res. Commun.* **354**, 591–597 (2007).
15. Tanaka, T. *et al.* Free fatty acids induce cholecystokinin secretion through *GPR120*. *Naunyn Schmiedebergs Arch. Pharmacol.* **377**, 523–527 (2008).
16. Miyauchi, S. *et al.* Distribution and regulation of protein expression of the free fatty acid receptor *GPR120*. *Naunyn Schmiedebergs Arch. Pharmacol.* **379**, 427–434 (2009).
17. Kido, Y. *et al.* Tissue-specific insulin resistance in mice with mutations in the insulin receptor, IRS-1, and IRS-2. *J. Clin. Invest.* **105**, 199–205 (2000).
18. Bernal-Mizrachi, E., Wen, W., Stahlhut, S., Welling, C. M. & Permutt, M. A. Islet beta cell expression of constitutively active *Akt1/PKB* alpha induces striking

19. Hosooka, T. *et al.* Dok1 mediates high-fat diet-induced adipocyte hypertrophy and obesity through modulation of PPAR- γ phosphorylation. *Nature Med.* **14**, 188–193 (2008).
20. Ntambi, J. M. *et al.* Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. *Proc. Natl. Acad. Sci. USA* **99**, 11482–11486 (2002).
21. Gutierrez-Juarez, R. *et al.* Critical role of stearoyl-CoA desaturase-1 (SCD1) in the onset of diet-induced hepatic insulin resistance. *J. Clin. Invest.* **116**, 1686–1695 (2006).
22. Jeyakumar, S. M. *et al.* Fatty acid desaturation index correlates with body mass and adiposity indices of obesity in Wistar NIN obese mutant rat strains WNIN/Ob and WNIN/GR-Ob. *Nutr. Metab. (Lond.)* **6**, 27 (2009).
23. Brown, J. M. *et al.* Combined therapy of dietary fish oil and stearoyl-CoA desaturase 1 inhibition prevents the metabolic syndrome and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **30**, 24–30 (2010).
24. Ichimura, A., Ruike, Y., Terasawa, K., Shimizu, K. & Tsujimoto, G. MicroRNA-34a inhibits cell proliferation by repressing mitogen-activated protein kinase kinase 1 during megakaryocytic differentiation of K562 cells. *Mol. Pharmacol.* **77**, 1016–1024 (2010).
25. Hara, T. *et al.* Novel selective ligands for free fatty acid receptors *GPR120* and *GPR40*. *Naunyn Schmiedebergs Arch. Pharmacol.* **380**, 247–255 (2009).
26. Sun, Q. *et al.* Structure-activity relationships of *GPR120* agonists based on a docking simulation. *Mol. Pharmacol.* **78**, 804–810 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to all subjects who participated in these studies. In Japan, the study was supported in part by research grants from the Japan Society for the Promotion of Science; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Japan Science and Technology Agency; and the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program), initiated by the Council for Science and Technology Policy. A.I. is a fellow of the Japan Society for the Promotion of Science. A.B. is a fellow of the EU-funded EUROCHIP consortium. In France, the study was supported by le Conseil Régional Nord Pas de Calais/FEDER and the Agence Nationale de la Recherche (Programme de Recherche en Nutrition et Alimentation, SensoFAT). The Northern Finland Birth Cohort Studies 1986 received financial support from the Academy of Finland, the University Hospital of Oulu (Finland), the University of Oulu (Finland), the European Commission (EURO-BLCS, Framework 5 award QLG1-CT-2000-01643), and the Medical Research Council (G0500539, G06000705, PrevMetSyn/SALVE). We thank the ABOS consortium and the CIC-CCPPRB (Lille CHRU) team for their help in sample handling and clinical data collection. We are grateful to M. Deweerder and F. Allegaert for human DNA bank management.

Author Contributions A.I., A. Hirasawa, O.P.-G. and A.B. are equally contributing first authors. G.T. and P.F. had the ideas for the mouse and human projects, respectively. A.I., A. Hirasawa, A.B., P.F. and G.T. drafted the manuscript. O.P.-G., H.C., D.M. and I.W. edited the manuscript and contributed to discussions. A. Hirasawa and G.T. designed the mouse research. A.I., A. Hirasawa, K.I. and G.T. created *Gpr120*-mutant mice. A.I., A. Hirasawa, A. Körner, T.H., I.K., T.-a.K., K.A., M. Takeuchi, K.O., N.L. and G.T. conducted biochemical and histochemical analyses for the mouse study. A.I. and A. Hirasawa performed bioinformatic analysis for the mouse study. L.Y. and C.L. performed the statistical analyses, and A.B. contributed to statistical analyses for the human study. O.P.-G. and I.W. designed the human expression gene study. A.L. performed the human expression gene study. H.C. and S.V. performed *GPR120* sequencing and variant genotyping, respectively. P.B., M. Tauber, C.M., A.M., R.B., P.E., M.-R.J., W.V.H., L.V.G., F.H., B.B., C.L.-M., K.R., A. Kouvatsi and F.P. contributed to cohort-study samples and researched data.

Author Information Microarray data have been deposited in the NCBI Gene Expression Omnibus under accession number GSE32095. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.F. (p.froguel@imperial.ac.uk) or G.T. (gtsuji@pharm.kyoto-u.ac.jp).

METHODS

Generation and genotyping of GPR120-deficient mice. GPR120-deficient mice on a mixed C57Bl/6/129 background were generated by homologous recombination. Exon 1 of the *Gpr120* gene was replaced with a PGK-neo cassette (Supplementary Fig. 1).

Animals. Mice were housed under a 12-hr light–dark cycle and given regular chow, MF (Oriental Yeast Co.). For HFD studies, 5-week-old male mice were placed on a 58Y1 diet (PMI Nutrition International) for a total period of 11 weeks. The methods used for animal care and experimental procedures were approved by the Animal Care Committee of Kyoto University.

Indirect calorimetry. Twenty-four-hour energy expenditure and respiratory quotient (RQ) were measured by indirect calorimetry, using an open-circuit calorimeter system (MK-5000RQ, Muromachi Kikai Co.). Respiratory quotient is the ratio of carbon dioxide production to oxygen consumption (VO_2). Energy expenditure was calculated as the product of the calorific value of oxygen ($3.815 + 1.232RQ$) and VO_2 . Locomotor activity was measured by using an infrared-ray passive sensor system (Supermex, Muromachi Kikai Co.).

Histology and immunohistochemistry. Epididymal adipose and pancreatic tissues were fixed in 10% neutral-buffered formalin, embedded in paraffin, and sectioned at 5 μ m. H&E staining was performed using standard techniques. To measure the diameter of adipocytes and the area of pancreatic islets, the diameters of 100 cells from five sections from each group were measured using NIH IMAGE software. More than 10 fields were examined, islet area was traced and total islet area was calculated and expressed as the average score. Liver tissues were embedded in OCT compound (Sakura Finetech) and snap-frozen in liquid nitrogen. Tissue sections were stained with Oil Red O (Sigma-Aldrich) for lipid deposition using standard methods.

Triglyceride content assay. To determine the triglyceride content of liver, tissue was homogenized with 1/2.5/1.25 (vol/vol) 0.5 M acetic acid/methanol/chloroform. The mixture was shaken and 1.25 volumes of chloroform added. The mixture was shaken overnight and then 1.25 volumes of 0.5 M acetic acid added. After centrifugation at 1,500g for 10 min, the organic layer was collected, dried and resuspended in 100% isopropyl alcohol. Measurements were conducted using Triglyceride E-test Wako (Wako).

Glucose tolerance and insulin tolerance tests. Glucose tolerance assays were performed on 24-hr-fasted mice. After baseline glucose values were individually established using One Touch Ultra (LifeScan), each mouse was given an intra-peritoneal injection of 1.5 mg glucose per gram of body weight. Insulin tolerance was conducted using the same glucometer. After baseline glucose values were established, mice were given human insulin (0.75 mU g⁻¹ intraperitoneal; Sigma-Aldrich). Clearance of plasma glucose was subsequently monitored at 15, 30, 60, 90 and 120 min post-injection.

Immunoblot analysis. For insulin stimulation, 5 U insulin (Sigma-Aldrich) was injected through the inferior vena cava. Five minutes later, samples of liver, skeletal muscle or WAT were dissected and immediately frozen in liquid nitrogen. Immunoblot analysis were performed as described previously^{5,24,25}. Anti-IRS1 (Millipore), anti-IRS2 (Millipore), anti-SCD1 (Santa Cruz Biotechnology), anti-IR β anti-AKT (Cell Signaling Technology), anti-p-AKT (Cell Signaling Technology) and anti- β -actin (Sigma-Aldrich) antibodies were used as the primary antibodies.

Mouse gene expression analysis. Total RNA was extracted from tissue or cells using ISOGEN (Nippon Gene). Quantitative RT-PCR and microarray analysis were performed as described previously^{4,26}. Briefly, genome-wide mRNA expression profiles were obtained by microarray analysis with the Affymetrix GeneChip Mouse 430 2.0 Array, according to the manufacturer's instructions. We used the robust multi-array analysis expression measure that represents the log-transformation of intensities (background corrected and normalized) from the GeneChips²⁷. Functional associations between differentially expressed genes were analysed using Ingenuity Pathways Analysis (version 4.0, Ingenuity Systems).

Microcomputed tomography scanning. Images were obtained using a micro-computed tomography system (SHIMADZU ClairvivoCT) with a high-resolution flat-panel detector. The maximum resolution of this modality was less than 40 μ m. The scanner was assumed to have a cylindrical field of view of 65.3 mm in section view and of 300 mm in transaxial view. The X-ray source was biased at 60 keV with the anode current set to 160 μ A. Computed tomography images were analysed with OSIRIX software (<http://www.osirix-viewer.com/>).

Fatty acid composition of epididymal WAT, liver and plasma. Esterified and non-esterified fatty acid composition was measured by gas chromatography. Briefly, to analyse esterified fatty acid, samples of epididymal adipose tissue (20–25 mg), liver (25–30 mg) and plasma (100 μ l) were snap-frozen in liquid nitrogen and homogenized in 4 ml of 0.5 N KOH-methanol. Samples were then boiled at 100 °C for 30 min to hydrolyse. Total lipids in each sample homogenate were then extracted with hexane, followed by trans-esterification of fatty acids using boron trifluoride-methanol at 100 °C for 15 min. Methylated fatty acids were then extracted with hexane and analysed using a GC-2010AF gas chromatograph

(SHIMADZU). For the analysis of non-esterified fatty acid, samples of epididymal adipose tissue (10–15 mg), liver (10–15 mg) and plasma (100 μ l) were snap-frozen in liquid nitrogen and homogenized in a mixture of 1.2 ml water, 3 ml methanol and 1.5 ml chloroform. Total lipids in each sample homogenate were extracted with a mixture of 1.2 ml of water and 1.2 ml of chloroform, followed by silylation of fatty acids using N,O-bis(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane at 100 °C for 60 min. Silylated fatty acids were then extracted with hexane and analysed using a GC-2010AF gas chromatograph (SHIMADZU). **Mouse embryonic fibroblast adipogenesis assay.** To prepare MEFs, we minced 13.5-d-post-coital mouse embryos and digested them with trypsin. Cells were collected and cultured in modified Eagle's medium (α -MEM; supplemented with 10% fetal bovine serum (FBS), 50 U ml⁻¹ penicillin and 50 mg ml⁻¹ streptomycin). We induced confluent MEFs to undergo adipogenic differentiation by incubating them first for 2 d with 10 μ g ml⁻¹ insulin, 250 nM dexamethasone and 0.5 mM isobutylmethylxanthine (Sigma-Aldrich). We measured cellular triglyceride content with Triglyceride E-test Wako (Wako).

Lipid infusion. Intralipid solution with 2 mM triglycerides:palmitoleate was prepared using a previously described protocol⁴. Briefly, lipids were dissolved in a solvent containing 5% glycerol and 0.72% phosphocholine in 0.9% saline and sonicated repeatedly. Lipids stayed in suspension for one week and had to be vortexed well before loading the syringe and tubing to prevent clogging. Before lipid infusion, mice were anaesthetized and an indwelling catheter was inserted in the left internal jugular vein. After overnight fasting, lipids were infused at a rate of 500 ml kg⁻¹ h⁻¹ for 6 h. At the end of the infusion, tissues were collected.

Statistical analysis of the GPR120-deficient mouse study. The level of significance for the difference between data sets was assessed using Student's *t*-test. Analysis of variance followed by Tukey's test was used for multiple comparisons. Data were expressed as mean \pm s.e.m. $P < 0.05$ was considered to be statistically significant.

Human study population. The study protocol was approved by all local ethics committees and informed consent was obtained from each subject before participation in the study, in accordance with the Declaration of Helsinki principles. For children younger than 18 years, an oral consent was obtained and parents provided written informed consent. All subjects were of European origin.

Human gene expression analysis. We used liver, subcutaneous and omental adipose tissue samples from the *Atlas Biologique de l'Obésité Sévère* (ABOS) cohort (ClinicalGov NCT01129297), a cohort studied in the Département de Chirurgie Générale et Endocrinienne²⁸ (Lille CHRU). Total RNA was extracted from the tissues using an RNeasy protect Mini Kit (QIAGEN) and quantified by absorbance at 260 nm and 280 nm in a PerkinElmer spectrophotometer. Human *GPR120*, *FABP4*, *PPARG*, *CD68* and *SCD* mRNA levels were quantified by reverse transcription reaction followed by qRT-PCR. Quantitative assessment of human mRNA expression was performed using TaqMan Gene Expression Assays (Hs0111664_m1: *GPR120* and Hs9999905_m1: *GAPDH*; Hs00609791_m1: *FABP4*; Hs00234592_m1: *PPARG*; Hs00154355_m1: *CD68*; Hs01682761_m1: *SCD*; Applied Biosystems) with an Applied Biosystems 7900HT Fast Real-Time PCR System. As an internal control for potential housekeeper reference variability, gene transcript levels were normalized to *GAPDH* reference housekeeper transcript level. The mean of the triplicate cycle thresholds of the target was normalized to the mean of triplicate cycle thresholds of the reference internal housekeeper genes using the formula $2^{CT_{GAPDH} - CT_{target}}$, which yielded a relative target-to-reference transcript concentration value as a fraction of reference transcript. Samples for which the cycle threshold was above 35 were excluded from the analysis.

GPR120 exon sequencing. We sequenced the four *GPR120* exons in 312 obese patients including 121 French obese adults and 191 French obese children who were recruited by the CNRS-UMR8199 unit and the Department of Nutrition of Paris Hotel Dieu Hospital. *GPR120* is located on human chromosome 10q23.33 and encodes a 377-amino-acid protein (NCBI NM_181745.3 and NP_859529). PCR conditions and primer sequences are available on request. Fragments were bidirectionally sequenced using the automated 3730xl DNA Analyzer (Applied Biosystems). Electrophoregram reads were assembled and analysed using the VARIANT REPORTER software (Applied Biosystems). The locations of the variants are displayed by base numbers counting from the ATG translation initiation codon following the Human Genome Variation Society nomenclature for the description of sequence variations. The positions of mutations were indicated by reference to the human genome build NCBI36/hg18.

Genotyping of the p.R270H and p.R67C/rs6186610 variants. We genotyped the two non-synonymous variants in 6,942 unrelated obese subjects and in 7,654 control subjects, all of European descent. Genotyped populations are described in Supplementary Table 4. The set of obese subjects included 516 unrelated French obese children who were recruited by the CNRS-UMR8199 unit or Toulouse Children's Hospital²⁹; 332 Italian obese children from Verona³⁰ or Rome³¹; 170 Finnish obese adolescents from the Northern Finland Birth Cohort 1986³².

(NFBC1986); 1,164 unrelated French obese adults from the ABOS cohort²⁸ or recruited by the CNRS-UMR8199 unit and the Department of Nutrition of Paris Hotel Dieu Hospital²⁹; 2,514 Belgian obese patients from the outpatient obesity clinic at Antwerp University Hospital³³; 1,736 Swiss obese subjects who were recruited after gastric surgery in Zurich³⁴; and 510 Greek obese subjects recruited in the Hippokration Hospital of Thessaloniki or in the Second Department of Internal Medicine of the Hospital of Alexandroupolis³⁵. The set of control subjects included 422 Italian lean children from Verona³⁰, 4,639 Finnish lean adolescents from the NFBC1986 cohort³²; 1,976 French lean adults from the D.E.S.I.R (Data from the Epidemiological Study on the Insulin Resistance syndrome) prospective study³⁶ and from the Haguenau study³⁷; 148 Belgian lean subjects from Antwerp Hospital³³; and 469 Greek lean individuals recruited in medical examination centres in Thessaloniki³⁵. The 1,109 French pedigrees selected for obesity were recruited by the CNRS-UMR8199³⁸, and the 780 German childhood obesity trios were recruited at the Universities of Marburg and Essen³⁹. The p.R270H variant was genotyped using the LightCycler 480 High Resolution Melting (HRM) Master kit (Roche), following the manufacturer's protocol. Genotyping of p.R67C/rs6186610 was performed using a custom TaqMan assay according to the manufacturer's instructions (Applied Biosystems). Allelic discrimination was performed using an Applied Biosystems 7900HT Fast Real-Time PCR System and SDS 2.3 software. For both variants, genotype success rate was at least 95% and no deviation ($P > 0.05$) from Hardy-Weinberg equilibrium was observed in any of the examined populations.

Phenotyping. The 90th and 97th body mass index (BMI) percentiles were used as thresholds for childhood overweight and obesity, respectively, according to the recommendations of the European Childhood Obesity Group study in local reference populations^{40,41}. Adult subjects were defined as normal ($BMI < 25 \text{ kg m}^{-2}$), overweight ($25 \leq BMI < 30 \text{ kg m}^{-2}$) and obese ($BMI \geq 30 \text{ kg m}^{-2}$) according to the International Obesity Task Force recommendations.

In silico analysis of both p.R270H and p.R67C variants. Phylogenetic conservation of the part of *GPR120* containing each non-synonymous variant was assessed using the UCSC genome browser (Vertebrate Multiz Alignment & Conservation), based on a phylogenetic hidden Markov model, phastCons⁴². To predict the possible effect of both amino-acid substitutions on the structure and function of GPR120, we used several programs: the POLYPHEN (polymorphism phenotyping) web-based program^{43,44}; PANTHER⁴⁵ (protein analysis through evolutionary relationships); the SIFT (sorting intolerant from tolerant) algorithm⁴⁶; the SNAP (screening for non-acceptable polymorphisms) software⁴⁷; and the PMUT web-based program⁴⁸.

Plasmid construction. A FLAG-human GPR120/pcDNA5/FRT/TO plasmid was constructed by ligating GPR120 complementary DNA into the multicloning site of the mammalian expression vector pcDNA5/FRT/TO (Invitrogen) with the amino-terminal FLAG tag. The point mutation for constructing the FLAG-human GPR120 p.R67C/pcDNA5/FRT/TO and FLAG-GPR120 p.R270H/pcDNA5/FRT/TO plasmids was carried out using the following primers: p.R67C (sense: 5'-gggtctgtggcgccggacggcc-3'; antisense: 5'-ggcgtcgccggacggcc-3') and p.R270H (sense: 5'-aggcaccaggatccacgttccaggcaggac-3'; anti-sense: 5'-gtcctgtggggcacgtggatctggc-3'). All constructs were confirmed by DNA sequencing.

Cell lines and cell culture. Flp-In T-REx-293 (T-REx 293) cells (Invitrogen) were used to develop inducible and stable cell lines expressing GPR120 (wild type, p.R270H or p.R67C). T-REx 293 cells were routinely cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 10% FBS, 100 $\mu\text{g ml}^{-1}$ Zeocin (Invitrogen) and 10 $\mu\text{g ml}^{-1}$ blasticidin S (Funakoshi). T-REx 293 cells were transfected with FLAG-GPR120 (wild type, p.R270H or p.R67C)/pcDNA5/FRT/TO using Lipofectamine reagent (Invitrogen) and selected with DMEM, which had been supplemented with 10% FBS, 10 $\mu\text{g ml}^{-1}$ blasticidin S and 100 $\mu\text{g ml}^{-1}$ hygromycin B (Gibco BRL). GPR120 protein expression was induced by adding 10 $\mu\text{g ml}^{-1}$ of doxycycline hyclate (Dox; Sigma) for 48 h. Human NCI-H716 cells were obtained from the American Type Culture Collection (Manassas). Cells were grown in suspension in Roswell Park Memorial Institute 1640 medium supplemented with 10% FBS, 100 IU ml^{-1} penicillin and 100 $\mu\text{g ml}^{-1}$ streptomycin.

[Ca²⁺]_i response analysis. Cells were seeded at a density of 2×10^5 cells per well on collagen-coated 96-well plates, incubated at 37 °C for 21 h and then incubated in Hanks' Balanced Salt Solution (HBSS, pH 7.4) containing Calcium Assay Kit Component A (Molecular Devices) for 1 h at 20 °C. ALA used in the fluorometric imaging plate reader (FLIPR, Molecular Devices) assay was dissolved in HBSS containing 1% DMSO and prepared in another set of 96-well plates. These plates were set on the FLIPR, and mobilization of [Ca²⁺]_i evoked by agonists was monitored.

Transfection. One million cells were seeded into a 3.5-cm-diameter dish before transfection. NCI-H716 cells were transfected with 5 μg of each plasmid using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. At 24 h post-transfection, transfection of each FLAG-tagged construct was confirmed by anti-FLAG FACS analysis. Then the cells were reseeded in 24-well culture plates

coated with Matrigel matrix (BD Biosciences) at a density of approximately 3×10^5 cells per well for the secretion studies. To test the effect of variant receptors on the ALA-induced [Ca²⁺]_i response of the wild-type receptor, 2 $\times 10^7$ T-REx 293 cells expressing Dox-inducible FLAG-GPR120 wild type were seeded into a 15-cm-diameter dish before transfection. Cells were then transfected with 32 μg of each plasmid (empty vector, wild type and p.R270H GPR120) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. At 24 h post-transfection, cells were reseeded at a density of 2×10^5 cells per well on collagen-coated 96-well plates and treated with 10 $\mu\text{g ml}^{-1}$ of Dox, and at 48 h post-transfection, ALA-induced [Ca²⁺]_i response was monitored.

GLP-1 secretion analysis. Cells were serum-starved with FBS-free DMEM for 3 h, washed with HBSS and incubated for 2 h at 37 °C in 0.3 ml FBS-free DMEM containing DMSO (negative control), 1 μM phorbol 12-myristate 13-acetate (positive control) or 100 μM ALA. Supernatants were collected and the active GLP-1 concentration in the supernatant was determined by enzyme immunoassay using GLP-1 (Active) ELISA Kit (Millipore).

Flow cytometry analysis. Anti-FLAG (Sigma) and anti-HA (Roche) antibodies were used for staining. Data were acquired and analysed on FACSCalibur with CELLQUEST software (Becton Dickinson).

Statistical analysis of human study. We assessed the effect of both non-synonymous variants (p.R270H and p.R67C) on obesity using a logistic regression adjusted first for age and gender and then for age, gender and geography, under an additive model, using the software R (version 2.12). Adjustment for geography was achieved to reflect a north-south gradient between the six different countries of origin of the study participants. An ordinal variable was created and coded: 1 for Finland, 2 for Belgium, 3 for France and Switzerland, 4 for Italy and 5 for Greece. This variable was added as a covariate in the logistic regression model.

Data analysis for the [Ca²⁺]_i response was performed using IGOR PRO (WaveMetrics). Significant differences between expression among wild-type and heterozygous groups, and among lean and obese wild-type subjects, were assessed using non-parametric Mann-Whitney analysis (GRAPHPAD PRISM 5 software).

27. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
28. Poulain-Godefroy, O., Lecoeur, C., Pattou, F., Fruhbeck, G. & Froguel, P. Inflammation is associated with a decrease of lipogenic factors in omental fat in women. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **295**, R1–R7 (2008).
29. Meyre, D. *et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nature Genet.* **41**, 157–159 (2009).
30. Morandi, A. *et al.* The Q121 variant of ENPP1 may protect from childhood overweight/obesity in the Italian population. *Obesity (Silver Spring)* **17**, 202–206 (2009).
31. Buzzetti, R. *et al.* PPAR- γ 2 Pro12Ala variant is associated with greater insulin sensitivity in childhood obesity. *Pediatr. Res.* **57**, 138–140 (2005).
32. Järvelin, M. R. *et al.* Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. *Paediatr. Perinat. Epidemiol.* **11**, 298–312 (1997).
33. Peeters, A. *et al.* Variants in the FTO gene are associated with common obesity in the Belgian population. *Mol. Genet. Metab.* **93**, 481–484 (2008).
34. Steffens, R., Potoczna, N., Bieri, N. & Horber, F. F. Successful multi-intervention treatment of severe obesity: a 7-year prospective study with 96% follow-up. *Obes. Surg.* **19**, 3–12 (2009).
35. Rouskas, K. *et al.* Association between BBS6/MKKS gene polymorphisms, obesity and metabolic syndrome in the Greek population. *Int. J. Obes. (Lond.)* **32**, 1618–1625 (2008).
36. Balkau, B. An epidemiologic survey from a network of French Health Examination Centres (D.E.S.I.R.): epidemiologic data on the insulin resistance syndrome [in French]. *Rev. Epidemiol. Sante Publique* **44**, 373–375 (1996).
37. Jaquet, D., Collin, D., Levy-Marchal, C. & Czernichow, P. Adult height distribution in subjects born small for gestational age. *Horm. Res.* **62**, 92–96 (2004).
38. Meyre, D. *et al.* R125W coding variant in TBC1D1 confers risk for familial obesity and contributes to linkage on chromosome 4p14 in the French population. *Hum. Mol. Genet.* **17**, 1798–1802 (2008).
39. Jarick, I. *et al.* Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum. Mol. Genet.* **20**, 840–852 (2011).
40. Poskitt, E. M. Defining childhood obesity: the relative body mass index (BMI). European Childhood Obesity group. *Acta Paediatr.* **84**, 961–963 (1995).
41. Rolland-Cachera, M. F. *et al.* Body mass index variations: centiles from birth to 87 years. *Eur. J. Clin. Nutr.* **45**, 13–21 (1991).
42. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
43. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894–3900 (2002).
44. Sunyaev, S. *et al.* Prediction of deleterious human alleles. *Hum. Mol. Genet.* **10**, 591–597 (2001).
45. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
46. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**, 1073–1081 (2009).

47. Bromberg, Y., Yachdav, G. & Rost, B. SNAP predicts effect of mutations on protein function. *Bioinformatics* **24**, 2397–2398 (2008).
48. Ferrer-Costa, C. et al. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* **21**, 3176–3178 (2005).

Enzymatic catalysis of anti-Baldwin ring closure in polyether biosynthesis

Kinya Hotta^{1*}, Xi Chen^{1*}, Robert S. Paton², Atsushi Minami³, Hao Li¹, Kunchithapadam Swaminathan¹, Irimpan I. Mathews⁴, Kenji Watanabe⁵, Hideaki Oikawa³, Kendall N. Houk⁶ & Chu-Young Kim¹

Polycyclic polyether natural products have fascinated chemists and biologists alike owing to their useful biological activity, highly complex structure and intriguing biosynthetic mechanisms. Following the original proposal for the polyepoxide origin of lasalocid and isolasalocid¹ and the experimental determination of the origins of the oxygen and carbon atoms of both lasalocid and monensin, a unified stereochemical model for the biosynthesis of polyether ionophore antibiotics was proposed². The model was based on a cascade of nucleophilic ring closures of postulated polyepoxide substrates generated by stereospecific oxidation of all-*trans* polyene polyketide intermediates². Shortly thereafter, a related model was proposed for the biogenesis of marine ladder toxins, involving a series of nominally disfavoured anti-Baldwin, *endo*-tet epoxide-ring-opening reactions^{3–5}. Recently, we identified Lsd19 from the *Streptomyces lasaliensis* gene cluster as the epoxide hydrolase responsible for the epoxide-opening cyclization of bisepoxyprelasalocid A⁶ to form lasalocid A^{7,8}. Here we report the X-ray crystal structure of Lsd19 in complex with its substrate and product analogue⁹ to provide the first atomic structure—to our knowledge—of a natural enzyme capable of catalysing the disfavoured epoxide-opening cyclic ether formation. On the basis of our structural and computational studies, we propose a general mechanism for the enzymatic catalysis of polyether natural product biosynthesis.

Epoxide-opening cascades, including those using the disfavoured anti-Baldwin cyclization (Fig. 1a), have emerged as an important synthetic strategy in the field of organic chemistry¹⁰. Whereas little is known about ladder polyether biosynthesis, biosynthesis of ionophore polyethers (Fig. 1b) has been studied extensively¹¹. Although most cyclic ethers in ionophore polyethers are the favoured *exo*-cyclization products (Fig. 1b, in blue), several compounds, including lasalocid A, contain six-membered rings presumably formed via a disfavoured *endo*-tet cyclization (Fig. 1b, in red).

Lsd19 was co-crystallized with the substrate analogue at pH 4.6, and the crystal structure was determined to 1.59 Å resolution (Supplementary Table 1). Lsd19 consists of two highly homologous domains (Fig. 2a) that belong to the nuclear transport factor 2 (NTF2)-like superfamily¹², which includes Δ^5 -3-ketosteroid isomerases (KSIs)¹³ and limonene-1,2-epoxide hydrolase (LEH)¹⁴. Each domain consists of three α -helices and a cone-like six-stranded β -sheet whose cavity forms a deep substrate-binding pocket. The amino- and carboxy-terminal domains (Lsd19A and Lsd19B, respectively) are linked by a short loop and are oriented in a head-to-tail fashion, resembling the homodimer structure of single-domain NTF2-like proteins¹⁵. Lsd19A and Lsd19B have identical backbone conformations; their respective 117 core $\text{C}\alpha$ atoms have a root-mean-squared deviation of just 0.79 Å. There are two significant differences between Lsd19A and Lsd19B. First, Lsd19B has an additional 18-residue-long loop–helix–loop near

the entrance to its substrate-binding pocket (Fig. 2a, in green). Second, the active site of Lsd19A contains an uncyclized bisepoxide substrate (Fig. 2b), whereas the active site of Lsd19B contains a doubly cyclized tetrahydrofuran–tetrahydrofuran (THF–THF) product. This difference in active sites is consistent with our previously reported results that indicated that Lsd19A catalyses a single round of cyclization of the bisepoxide substrate, whereas Lsd19B cyclizes the initially formed THF ether/monoepoxide intermediate, predominantly to a tetrahydrofuran–tetrahydropyran (THF–THP) product¹¹.

In Lsd19A, the uncyclized substrate analogue is seen in the active site. This observation is consistent with the fact that Lsd19 is inhibited by the low pH of the crystallization buffer (Supplementary Fig. 1a). This characteristic is shared by many EHs^{15,16}. The electron density for

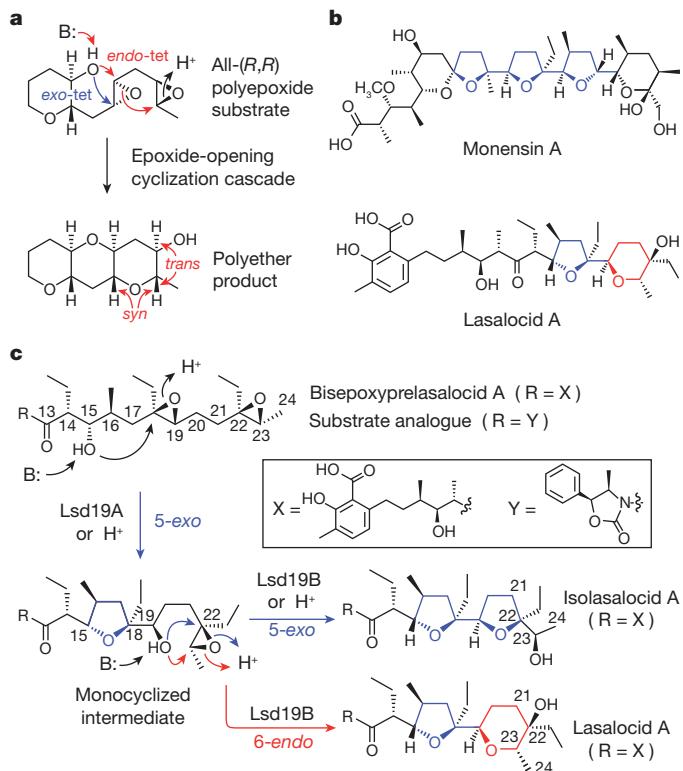


Figure 1 | Polyether natural products and proposed steps in the cyclic ether formation. **a**, An example of a cascade epoxide ring closure implicated in the biosynthesis of ladder polyether. **b**, Structures of representative ionophore polyethers. **c**, Plausible mechanism of epoxide-mediated cyclic ether formation leading to the formation of lasalocid A, isolasalocid A and the product analogues.

¹National University of Singapore, Department of Biological Sciences, 14 Science Drive 4, 117543 Singapore. ²Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK. ³Hokkaido University, Division of Chemistry, Graduate School of Science, North 10 West 8, Sapporo 060-0810, Japan. ⁴Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, 2575 Sand Hill Road MS 99, Menlo Park, California 95124, USA. ⁵University of Shizuoka, Division of Pharmaceutical Sciences, Graduate School of Pharmaceutical Sciences, 52-1 Yada, Suruga-ku, Shizuoka City, Shizuoka 422-8526, Japan. ⁶University of California Los Angeles, Department of Chemistry and Biochemistry, 607 Charles E. Young Drive East, Box 951569, Los Angeles, California 90095-1569, USA. *These authors contributed equally to this work.

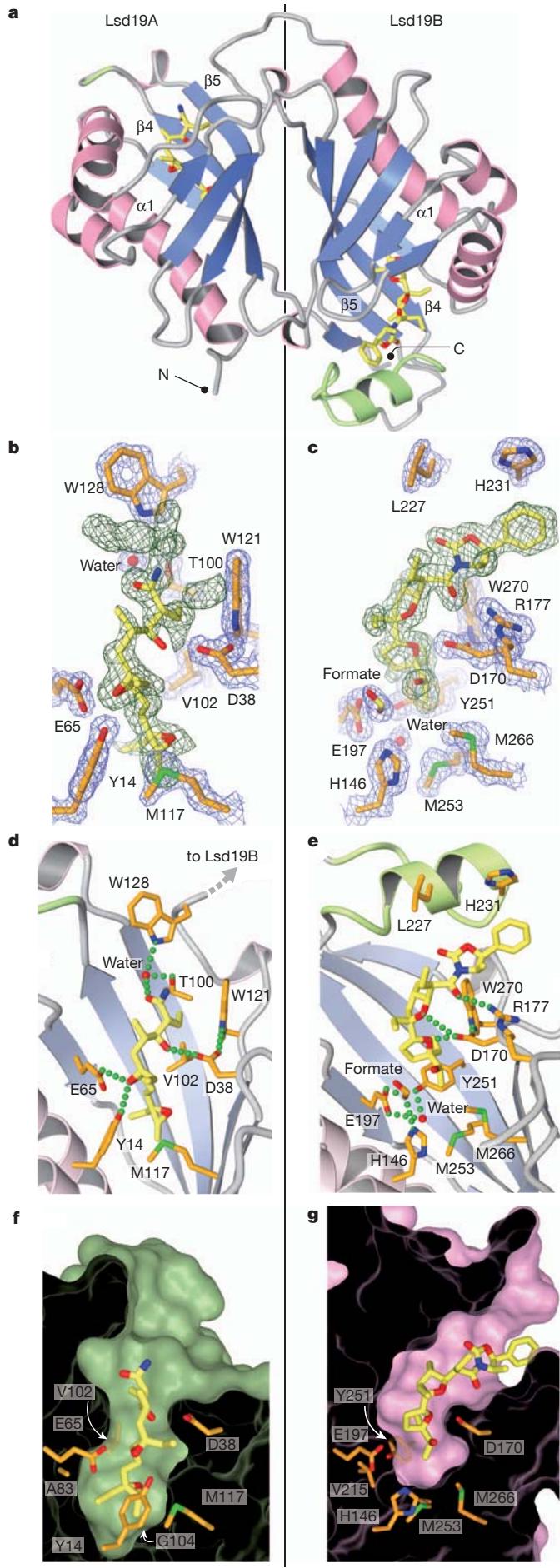


Figure 2 | Crystal structure of Lsd19. **a**, The overall fold of Lsd19. The extra loop–helix–loop in Lsd19B and the corresponding insertion site in Lsd19A are shown in green. **b, c**, Lsd19A (**b**) and Lsd19B (**c**) electron density for the ligand (simulated-annealing omit map contoured at 2.0σ , green mesh) and protein side chains ($2F_o - F_c$ map contoured at 1.5σ , blue mesh). Carbon atoms of the bound ligands and protein are shown in yellow and orange, respectively. Oxygen, nitrogen and sulphur atoms are shown in red, blue and green, respectively. Green broken lines represent hydrogen bonds. **d, e**, Lsd19A (**d**) and Lsd19B (**e**) active-site hydrogen-bonding interactions between the ligand and protein side chains. The $\alpha 1$ helix (as labelled in panel **a**) was removed to clarify the view. **f, g**, Lsd19A (**f**) and Lsd19B (**g**) molecular surface cross-section, showing residues that determine the pocket size and shape.

the oxazolidinone portion of the substrate analogue is ill-defined and therefore omitted from the final model (Fig. 2b). As the oxazolidinone is not present in the natural substrate, disorder in this region is not unexpected. In contrast, the portion of the substrate that contains the two epoxide groups that eventually undergo cyclization is clearly defined in the electron density map (Fig. 2b). The terminal 22,23-epoxide sits at the bottom of the pocket and is surrounded by hydrophobic residues, protected from potential non-specific nucleophilic attack. The putative nucleophilic 15-hydroxyl oxygen is hydrogen bonded to Asp 38, whereas the oxygen atom of the electrophilic 18,19-epoxide that undergoes ring opening lies within hydrogen-bonding distance to Tyr 14 and Glu 65. Arg 54 is hydrogen-bonded to the catalytic Asp 38 via a water molecule, presumably promoting deprotonation and favourable positioning of the side-chain carboxylic acid of Asp 38 for ensuing proton abstraction from the 15-hydroxyl group. This active-site configuration is homologous to that of KSI¹³ and LEH¹⁴, which perform similar acid/base catalysis (Fig. 3b). Unlike KSI or LEH, however, Lsd19 must perform conformationally sensitive intramolecular cyclizations on a flexible substrate. Additional interactions between Lsd19A and the 3-methylsalicylate side of the native substrate can help fix the ligand conformation for the ensuing cyclization. Here, a water molecule hydrogen bonded by Thr 100 and Trp 128 forms a hydrogen bond with the carbonyl oxygen of the bound substrate analogue (Fig. 2d). An equivalent oxygen atom is present in the natural substrate, bisepoxyprelasalocid A.

In Lsd19B, an isolasalocid-like THF–THF compound is seen in the active-site pocket instead of the expected lasalocid-like THF–THP reaction product. This unexpected reaction compound is presumably a consequence of the acidic crystallization condition (Supplementary Fig. 1b). We have identified His 146, Asp 170, Glu 197 and Tyr 251 as primary candidates for the catalytic residues in Lsd19B (Fig. 2e). Like in Lsd19A, a hydrogen bond between Asp 170 and His 186 could raise the pK_a of His 186. The 13-carbonyl oxygen is hydrogen bonded to Arg 177, providing an additional enzyme–substrate interaction. The binding pocket of Lsd19B is shallower than that of Lsd19A due to the presence of Met 253 at the bottom of the pocket (Fig. 2g). This allows only the singly cyclized monoepoxide to bind in a catalytically relevant orientation, and promotes the critical substrate selectivity in these highly structurally similar domains. Although the shallower binding pocket in Lsd19B affords less substrate-binding energy available for confining the flexible substrate into the catalytically relevant conformation¹⁷, the additional loop–helix–loop structure (Fig. 2a, e, in green) seems to compensate for the lost substrate-binding surface.

Analysis of known polyether biosynthetic gene clusters reveals that two copies of polyether epoxide hydrolase (PEH)-encoding genes are present within each cluster: *monBI*, *monBII* (monensin)¹⁸ and *nigBI*, *nigBII* (nigericin)¹⁹. For *lsd19* (lasalocid)^{6,7} and *nanI* (nanchangmycin)²⁰, two copies are encoded as a single gene. The tetrodromycin gene cluster²¹ carries a single copy of PEH gene *tmnB*, presumably because tetrodromycin contains only one epoxide-derived THF ring. Sequence alignment of PEHs indicates that the catalytic residues identified in Lsd19 are highly conserved (Supplementary Fig. 2). This alignment also reveals that PEHs can be divided into two groups: PEH-A (Lsd19A-like) and PEH-B (Lsd19B-like). Only PEH-B has a bulky residue at

position 104 that makes the binding pocket shallower, and only PEH-B contains a loop–helix–loop insertion that provides additional ligand-binding surface.

To extend the structural knowledge gained from Lsd19, homology models of other PEHs were constructed. In these models, a direct correlation is observed between the depth of the binding pocket and the length of the substrate chain (Supplementary Fig. 3), leading us to propose a general mode of cyclic ether formation in polyether biosynthesis. With the deeper binding pockets, PEH-As are probably responsible for the formation of internal cyclic ethers. Interestingly, PEH-As that act on longer substrates, such as nigericin, carry extra PEH-A-specific C-terminal residues (Supplementary Fig. 2) positioned at the opening of the binding pocket (Supplementary Fig. 3a), possibly providing additional binding surface for the longer substrates. In contrast, PEH-Bs with shallower binding pockets probably catalyse the formation of terminal cyclic ethers. As seen in Lsd19B, the extra PEH-B-specific loop–helix–loop domain can provide the binding surface for the portion of the substrate protruding out from the shallow substrate-binding pocket.

As the current Lsd19B structure does not address the formation of THP directly, we used computational studies to understand how this enzyme catalyses the disfavoured 6-*endo* cyclization (Methods and Supplementary Tables 2–4). Quantum mechanical density functional calculations were performed to locate each transition structure for both 5-*exo* and 6-*endo* epoxide openings^{22,23} under acid- and base-catalysed conditions using model substrates, and to calculate the energy difference between the two processes (Fig. 3a and Supplementary Table 2). As expected²³, 5-*exo*-tet cyclization is favoured with both acid and base catalysis. However, the preference for the five-membered ring is reduced under base-catalysed conditions, due to the increased steric hindrance of nucleophilic attack at the more substituted epoxide terminus, as well as

the more product-like transition structure with shorter forming O–C bond lengths. The six-membered product is indeed thermodynamically favoured. This provides an indication of how Lsd19B may achieve the six-membered ring formation. We then explored how Lsd19B would influence the regioselectivity of the ring-opening transition state. Docking studies on the Lsd19B structure indicated that Asp 170, Glu 197 and Tyr 251 maintain close contact with the bound ligand. Similarities between the active site organization of Lsd19B and KSI¹³ (Fig. 3b) further support the idea that Asp 170 acts as a general base, whereas Glu 197 and Tyr 251 act as general acids stabilizing the developing transition structure oxyanion. This arrangement is also remarkably similar to the ‘theozyme’ derived from computations for the anti-Baldwin-cyclization antibody²³, which was eventually found to be closely related to the antibody structure²⁴. To test this idea quantitatively, competing 5-*exo* and 6-*endo* transition structures were optimized in the presence of Asp 170 and Tyr 251, whose geometry was constrained to the crystal structure coordinates. From this model, values of the activation energy and Gibbs free energy changes for ring closure were computed. These predict a preference for the 6-*endo* pathway of 2.5 kcal mol^{−1}, enough to achieve nearly 100:1 selectivity (Fig. 3c). The positioning of the catalytic general acid and general base groups causes the 6-*endo* transition state to be favoured, leading to the thermodynamically more stable THP product.

The structural knowledge gained here has provided a generalized view of the biosynthetic pathway involved in producing polyether natural products using only a pair of epoxide hydrolases. Furthermore, computational studies have provided an explanation for the regiochemical preference for the Lsd19-catalysed cyclization. Active-site pre-organization and general-base catalysis provides enzymatic control to overcome otherwise disfavoured chemical transformations. Similarly, stereochemically complex polyether ladder structures can be generated from

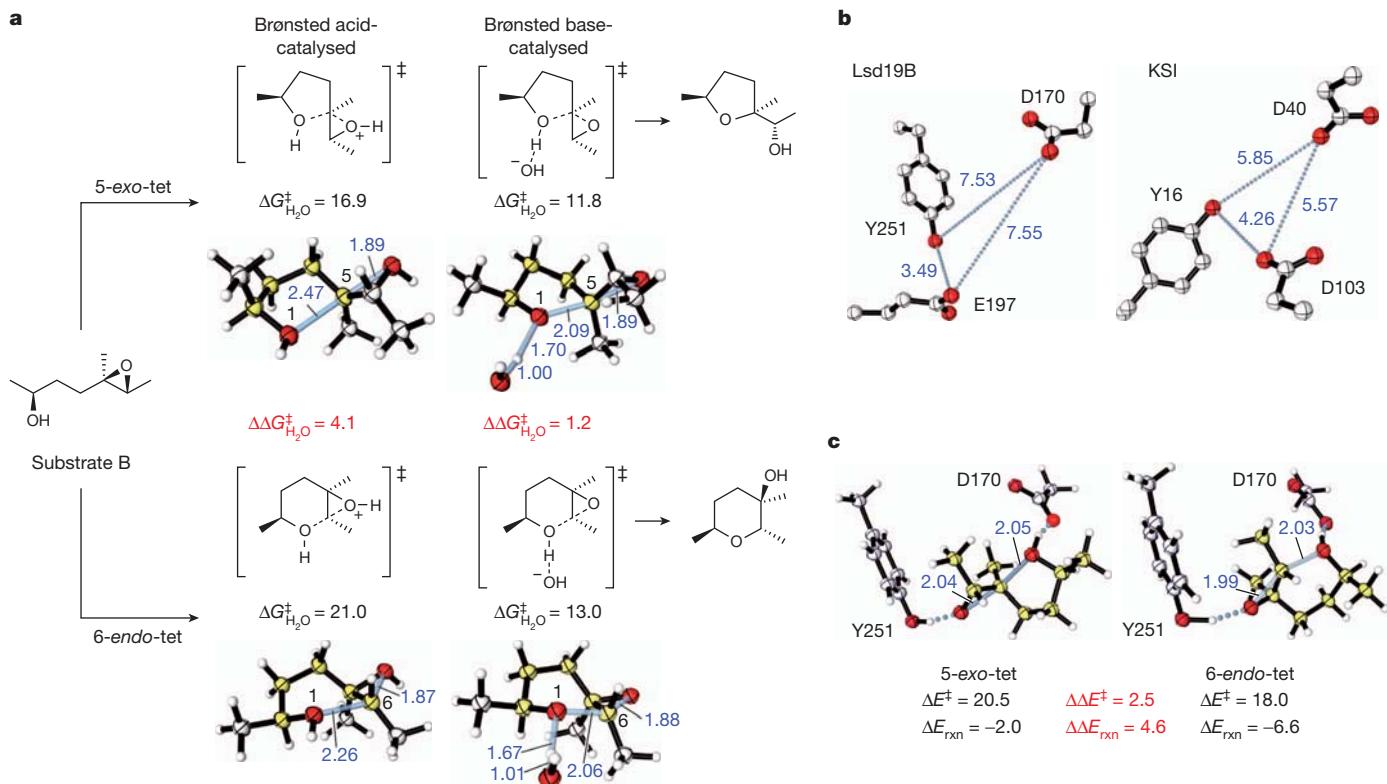


Figure 3 | Computational studies of the Lsd19-catalysed epoxide-opening cyclization reactions. **a**, Acid- versus base-catalysed cyclization of a model system via 5-*exo* or 6-*endo* cyclization, and the respective calculated lowest-energy competing transition structures. B2PLYP/6-311++G(d,p)//B2PLYP/6-31G(d) activation free energies (kcal mol^{−1}) and forming/breaking bond distances (in Ångstroms, shown in blue) are given. **b**, Comparison of active-site

geometries in Lsd19B and KSI (PDB accession 1E3V²⁴). **c**, ‘Theozyme’ calculations, optimizing the competing 5-*exo*-tet and 6-*endo*-tet transition structures surrounded by fixed catalytic residues. Relative energies obtained by B2PLYP/6-311++G(d,p)/M06-2X/6-31G(d) (kcal mol^{−1}) and forming/breaking bond distances are given.

a simple polyepoxide substrate by Lsd19B-like EHs that can form templating THP unit(s) that facilitate the subsequent cascade of step-wise *endo*-tet-selective epoxide-opening cyclizations.

METHODS SUMMARY

Lsd19 was produced using a variation of the published procedure⁸. Lsd19 was purified using metal affinity, anion exchange and size-exclusion chromatography steps to >95% purity and yielding approximately 2 mg protein per litre of culture. Initial crystals of Lsd19 with the presence of the substrate analogue⁸ were obtained from the Hampton Crystal Screen. Subsequent optimization resulted in a crystal that diffracted to 1.59 Å resolution at the Stanford Synchrotron Radiation Lightsource. Data were indexed and integrated using XDS²⁵. The structure of Lsd19 was determined by molecular replacement and refined by CNS²⁶ and REFMAC5 (CCP4)²⁷. Computational studies on the reaction mechanism were performed with B2PLYP/6-311++G(d,p) calculated energies on B2PLYP/6-31G(d) or M06-2X/6-31G(d) optimized geometries using the Gaussian 09 program²⁸. Docking Studies were performed with Autodock Vina²⁹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 May 2011; accepted 13 January 2012.

Published online 4 March 2012.

1. Westley, J. W., Blount, J. F., Evans, R. H. Jr, Stempel, A. & Berger, J. Biosynthesis of lasalocid. II. X-ray analysis of a naturally occurring isomer of lasalocid. *A. J. Antibiot. (Tokyo)* **27**, 597–604 (1974).
2. Cane, D. E., Celmer, W. D. & Westley, J. W. Unified stereochemical model of polyether antibiotic structure and biogenesis. *J. Am. Chem. Soc.* **105**, 3594–3600 (1983).
3. Nakanishi, K. The chemistry of brevetoxins: a review. *Toxicon* **23**, 473–479 (1985).
4. Baldwin, J. E. Rules for ring closure. *J. Chem. Soc. Chem. Comm.* 734–736 (1976).
5. Gilmore, K. & Alabugin, I. V. Cyclizations of alkynes: revisiting Baldwin's rules for ring closure. *Chem. Rev.* **111**, 6513–6556 (2011).
6. Shichijo, Y. et al. Epoxide hydrolase Lsd19 for polyether formation in the biosynthesis of lasalocid A: direct experimental evidence on polyene-polyepoxide hypothesis in polyether biosynthesis. *J. Am. Chem. Soc.* **130**, 12230–12231 (2008).
7. Migita, A. et al. Identification of a gene cluster of polyether antibiotic lasalocid from *Streptomyces lasaliensis*. *Biosci. Biotechnol. Biochem.* **73**, 169–176 (2009).
8. Smith, L., Hong, H., Spencer, J. B. & Leadlay, P. F. Analysis of specific mutants in the lasalocid gene cluster: evidence for enzymatic catalysis of a disfavoured polyether ring closure. *ChemBioChem* **9**, 2967–2975 (2008).
9. Minami, A. et al. Enzymatic epoxide-opening cascades catalyzed by a pair of epoxide hydrolases in the ionophore polyether biosynthesis. *Org. Lett.* **13**, 1638–1641 (2011).
10. Nakata, T. Total synthesis of marine polycyclic ethers. *Chem. Rev.* **105**, 4314–4347 (2005).
11. Westley, J. W. Antibiotic structure and biosynthesis. *J. Nat. Prod.* **49**, 35–47 (1986).
12. Marchler-Bauer, A. et al. CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.* **37**, D205–D210 (2009).
13. Ha, N. C., Choi, G., Choi, K. Y. & Oh, B. H. Structure and enzymology of Δ^5 -3-ketosteroid isomerase. *Curr. Opin. Struct. Biol.* **11**, 674–678 (2001).
14. Hopmann, K. H., Hallberg, B. M. & Himo, F. Catalytic mechanism of limonene epoxide hydrolase, a theoretical study. *J. Am. Chem. Soc.* **127**, 14339–14347 (2005).
15. van der Werf, M. J., Overkamp, K. M. & de Bont, J. A. Limonene-1,2-epoxide hydrolase from *Rhodococcus erythropolis* DCL14 belongs to a novel class of epoxide hydrolases. *J. Bacteriol.* **180**, 5052–5057 (1998).
16. Touhara, K. & Prestwich, G. D. Juvenile hormone epoxide hydrolase. Photoaffinity labeling, purification, and characterization from tobacco hornworm eggs. *J. Biol. Chem.* **268**, 19604–19609 (1993).
17. Jencks, W. P. Binding energy, specificity, and enzymic catalysis: the Circe effect. *Adv. Enzymol.* **43**, 219–410 (1975).
18. Leadlay, P. F. et al. Engineering of complex polyketide biosynthesis—insights from sequencing of the monensin biosynthetic gene cluster. *J. Ind. Microbiol. Biotechnol.* **27**, 360–367 (2001).
19. Harvey, B. M. et al. Insights into polyether biosynthesis from analysis of the nigericin biosynthetic gene cluster in *Streptomyces* sp. DSM4137. *Chem. Biol.* **14**, 703–714 (2007).
20. Sun, Y. et al. A complete gene cluster from *Streptomyces nanchangensis* NS3226 encoding biosynthesis of the polyether ionophore nanchangmycin. *Chem. Biol.* **10**, 431–441 (2003).
21. Demydchuk, Y. et al. Analysis of the tetrodomyycin gene cluster: insights into the biosynthesis of a polyether tetrotrate antibiotic. *ChemBioChem* **9**, 1136–1145 (2008).
22. Gruber, K. et al. Structural basis for antibody catalysis of a disfavored ring closure reaction. *Biochemistry* **38**, 7062–7074 (1999).
23. Na, J., Houk, K. N., Shevlin, C. G., Janda, K. D. & Lerner, R. A. The energetic advantage of 5-exo versus 6-endo epoxide openings: a preference overwhelmed by antibody catalysis. *J. Am. Chem. Soc.* **115**, 8453–8454 (1993).
24. Ha, N. C., Kim, M. S., Lee, W., Choi, K. Y. & Oh, B. H. Detection of large pK_a perturbations of an inhibitor and a catalytic group at an enzyme active site, a mechanistic basis for catalytic power of many enzymes. *J. Biol. Chem.* **275**, 41100–41106 (2000).
25. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
26. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
27. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
28. Frisch, M. J. et al. Gaussian 09, Revision A.2 (Gaussian, 2009).
29. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Royal Commission for the Exhibition of 1851 and Fulbright-AstraZeneca Research Fellowship (R.S.P.), the Japan Society for the Promotion of Science (No. LS103) (K.W.), the National Institutes of Health grant GM075962 (K.N.H.), the MEXT research grant on innovative area 22108002 (H.O.), and the National University of Singapore Life Sciences Institute Young Investigator Award (C.-Y.K.). Data collection was performed at the Stanford Synchrotron Radiation Lightsource. We thank D. W. Christianson, D. Hilvert and C. Khosla for critical reading and discussion of the manuscript.

Author Contributions A.M. and H.O. prepared the substrate analogue. K.W. cloned and purified Lsd19. X.C. and H.L. purified and crystallized Lsd19. X.C. and I.I.M. collected diffraction data and determined the structure. K.H., X.C. and I.I.M. refined the structure. K.S. provided assistance for crystallography. K.N.H. prepared and analysed models of Lsd19 homologues. R.S.P. and K.N.H. performed the computational study. C.-Y.K. conceived and supervised the project. C.-Y.K. prepared the manuscript with contributions from all co-authors.

Author Information The coordinates and structure factors have been deposited with the Protein Data Bank under accession number 3RGA. Reprints and permissions information is available at www.nature.com/reprints. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.-Y.K. (chuyoung@nus.edu.sg).

METHODS

Lsd19 expression and purification. Production of Lsd19 was performed partly according to the procedure described elsewhere⁸. Briefly, pCold-based pKW620 carrying the *lsd19* gene and the chaperone-encoding vector pG-KJE8 (Takara Bio) were introduced into *Escherichia coli* BL21(DE3) strain. The culture was grown in Luria Broth medium to OD_{600 nm} of 0.6, and the production of N-terminal His-tagged Lsd19 was induced by 100 μM isopropyl-β-D-galactoside, 0.5 mg ml⁻¹ L-arabinose and 5 ng ml⁻¹ tetracycline. The culture was incubated for another 20 h at 15 °C. Cells were harvested by centrifugation, resuspended in 50 mM sodium phosphate pH 7.4, 300 mM sodium chloride, 10 mM imidazole, 10% (v/v) glycerol and lysed by sonication. After centrifugation at 15,000 g for 45 min, the cleared supernatant was mixed with cobalt-agarose beads (Thermo Fisher Scientific). The mixture was incubated for 1 h at 4 °C with moderate shaking. After incubation, the mixture was loaded onto a column and washed with a wash buffer containing 50 mM sodium phosphate pH 7.4, 300 mM sodium chloride, 10 mM imidazole and 10% (v/v) glycerol. Lsd19 was eluted with a wash buffer supplemented with 150 mM imidazole. Fractions containing Lsd19 were pooled and exchanged into a buffer containing 20 mM Tris pH 8.5, 1 mM EDTA and 15 mM β-mercaptoethanol. Lsd19 was further purified by anion exchange chromatography using a HiTrapQ XL column followed by gel filtration on a Superdex200 10/300 GL column (GE Healthcare Life Sciences) in 50 mM potassium phosphate pH 7.0, 1 mM EDTA and 15 mM β-mercaptoethanol. The final protein was >95% pure as judged by polyacrylamide gel electrophoresis, and the yield was approximately 2 mg per litre of culture. The sample was further concentrated to 12 mg ml⁻¹ for storage.

Crystallization, data collection and structure determination. Crystals were obtained by the hanging-drop vapour diffusion method. The bisepoxide substrate analogue was prepared as described⁸. After combining Lsd19 (6 mg ml⁻¹) with the substrate analogue (100 mM in methanol) at 12.5:1 ratio (v/v) on ice, the sample was mixed with an equal volume of 0.1 M sodium acetate pH 4.6 and 1.6 M sodium formate and equilibrated at 18 °C. Crystals were further optimized using an additive 1,5-diaminopentane dihydrochloride in the presence of 200–700 μl Al's oil in the well. Before flash freezing, Lsd19 crystals were transferred for about 1 min to a cryoprotectant solution comprised of the crystallization buffer supplemented with 35% glycerol. X-ray diffraction data were collected from a single crystal to a 1.59 Å resolution at beam line 7-1 of Stanford Synchrotron Radiation Lightsource (SSRL; SLAC National Accelerator Laboratory). Data collection was performed on a single crystal at 100 K using a monochromatic X-ray at a wavelength of 0.9795 Å. Data were indexed and integrated using XDS²⁵. Structure factor was obtained by CTRUNCATE³⁰. The structure of Lsd19 was determined by molecular replacement with the program MOLREP³¹ using the *Pseudomonas putida* biotype B KSI structure (PDB accession 1DMN)³⁷ as a search model. The model was built by Coot³² and ARP/wARP³³ and refined by CNS²⁶ and REFMAC5 (CCP4)²⁷. Ramachandran plot of the refined model indicated that all of the residues assumed allowed backbone conformation with 98.2% within the favoured conformation. Other relevant data and refinement statistics are given in Supplementary Table 1.

Computation. Quantum chemical calculations were performed using Gaussian 09, revision A.2²⁸. All geometries were fully optimized. Transition structures were identified by a single imaginary harmonic vibrational frequency. Optimizations were performed with default convergence criteria at the B2PLYP/6-31G(d) level of theory using a fine grid for numerical integration. Free energies were computed in

the harmonic approximation at 298 K and 1 atm, with unscaled zero-point vibrational energies. Docking Studies were performed with Autodock Vina²⁹.

Single point energy calculations were performed to ensure basis set convergence at the B2PLYP/6-311++G(d,p)//B2PLYP/6-31G(d) level. Calculations were performed at the 'double-hybrid' density functional level of theory with the B2PLYP functional³⁴. This method replaces a fraction of the semi-local correlation energy by a non-local correlation energy expression that uses the Kohn–Sham orbitals in second-order perturbation theory and delivers improved energetics over hybrid density functionals. Results obtained at this level were also compared with the hybrid meta-generalized gradient approximation M06-2X density functional. B2PLYP and M06-2X calculations are in accord, both qualitatively and quantitatively, over the preference for 5-*exo*-tet over 6-*endo*-tet in the acid- versus base-catalysed cyclization of the model systems. Absolute barriers showed some variations; however, the energy differences between the two cyclization modes were consistent. Changing between double and triple zeta-valence polarized basis sets showed only small effects (Supplementary Table 2).

The effects of solvation were treated with single point energy calculations on the gas-phase geometries using a conductor-like polarizable continuum (CPCM) model³⁵ of water ($\rho = 78.36$) and dichloromethane ($\rho = 8.93$), defining the solute cavity by UFF radii. As for gas-phase values, B2PLYP and M06-2X results are in accord over preferred reaction pathway (Supplementary Table 3).

'Theozyme' calculations were performed using a truncated model of the active site, constraining the positions of the non-hydrogen atoms in the catalytic residues Glu 197 and Tyr 251 in their crystal structure coordinates. All other atom positions are optimized. Electrostatic effects of the protein interior were described with a CPCM model of low dielectric constant, using diethyl ether as solvent ($\rho = 4.24$)³⁶. A variety of methods show consistent energy differences between 5-*exo* and 6-*endo* pathways (Supplementary Table 4). Absolute energies and Cartesian coordinates for all species in the text are given in the Supplementary Data. Absolute gas-phase energies and free energies (1 atm, 298 K) from B2PLYP/6-31G(d) optimization, B2PLYP/6-311++G(d,p) single point energies, single point energies including CPCM solvation (UAKS solute radii) are all given in kcal mol⁻¹. Imaginary harmonic vibrations are in cm⁻¹.

30. Padilla, J. E. & Yeates, T. O. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D* **59**, 1124–1130 (2003).
31. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
32. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
33. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* **3**, 1171–1179 (2008).
34. Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **124**, 034108 (2006).
35. Barone, V. & Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **102**, 1995–2001 (1998).
36. Zhang, X. *et al.* Quantum mechanical design of enzyme active sites. *J. Org. Chem.* **73**, 889–899 (2008).
37. Kim, D. H. *et al.* Contribution of the hydrogen-bond network involving a tyrosine triad in the active site to the structure and function of a highly proficient ketosteroid isomerase from *Pseudomonas putida* biotype B. *Biochemistry* **39**, 4581–4589 (2000).

An oxygenase that forms and deoxygenates toxic epoxide

Robin Teufel¹, Thorsten Friedrich² & Georg Fuchs¹

Catabolism may give rise to toxic intermediates that compromise cell vitality, such as epoxide formation in the recently elucidated and apparently universal bacterial coenzyme A (CoA)-dependent degradation of phenylacetic acid¹. This compound is central to the catabolism of a variety of aromatics, such as phenylalanine, lignin-related compounds or environmental contaminants^{2,3}. The key phenylacetyl-CoA monooxygenase (epoxidase) of the pathway, PaaABCE^{1,4,5}, is also connected to the production of various primary and secondary metabolites^{6–9}, as well as to the virulence of certain pathogens^{1,10,11}. However, the enzyme complex has so far not been investigated in detail. Here we characterize the bacterial multi-component monooxygenase PaaABCE that, surprisingly, not only transforms phenylacetyl-CoA into its ring-1,2-epoxide, but also mediates the NADPH-dependent removal of the epoxide oxygen, regenerating phenylacetyl-CoA with formation of water. We provide evidence for a catalytic di-iron centre that is probably the key to the unprecedented deoxygenation of an organic compound by an oxygenase. Presumably, the bifunctionality is vital to avoid toxic intracellular epoxide levels if the subsequent catabolic steps are impeded. Our data suggest that detoxification is assisted by two thioesterases (PaaI and PaaY) forming non-reactive breakdown products. Hence, PaaABCE may harbour an intrinsic escape mechanism from its own toxic product and represents the archetype of a bifunctional oxygenase/deoxygenase. Analogous reactions may possibly be catalysed by other di-iron epoxidases.

PaaABCE catalyses a crucial step in bacterial degradation of phenylacetic acid (paa), namely, the epoxidation of phenylacetyl (pa)-CoA (Fig. 1, Supplementary Fig. 1). Its encoding genes are present in ~16%

of completely sequenced bacterial genomes¹. Among Archaea, it was mainly Halobacteria that acquired bacterial PaaABCE (Supplementary Fig. 2), as this group seemed inclined to take up bacterial genes¹². This enzyme is probably the most widespread member of bacterial multi-component monooxygenases that generally consist of oxygenase subunits, a small effector protein, and a reductase component¹³. According to sequence comparison, subunit A (PaaA) is the catalytic oxygenase compound harbouring conserved EX₂H-motives for iron binding and is a member of the class I di-iron proteins^{1,4,5} that are known to catalyse diverse and functionally unrelated enzymatic reactions, such as methane monooxygenation or ribonucleotide reduction¹⁴. PaaE resembles reductases with [2Fe-2S] and flavin cofactor binding domains; PaaE probably transfers electrons from NADPH to the di-iron centre of subunit A^{1,4,5}. Together, subunits PaaAE resemble benzoyl-CoA epoxidase BoxAB, which similarly transforms benzoyl-CoA into its ring-2,3-epoxide^{15,16}.

We determined cofactors, metal content, and acid-labile sulphur content of purified recombinant PaaABCE from *Pseudomonas* sp. strain Y2, and we also recorded ultraviolet-visible and electron paramagnetic resonance (EPR) spectra (Supplementary Figs 3–5). The composition of the complex was PaaA₂B_{3–4}C₂E₁ or multiples of this. Subunits B and C appear not to contain metals or cofactors, whereas a possible subunit PaaD did not copurify and had no effect on *in vitro* activity. It may be involved in maturation of PaaA or PaaE rather than catalysis^{1,4} (see discussion in Supplementary Information). We found 6.3 ± 0.1 mol Fe per mol of PaaA₂E₁, in good accordance with the presence of one [2Fe-2S] cluster per PaaE and one di-iron centre per PaaA, as further corroborated by the EPR spectra (Supplementary Fig. 6).

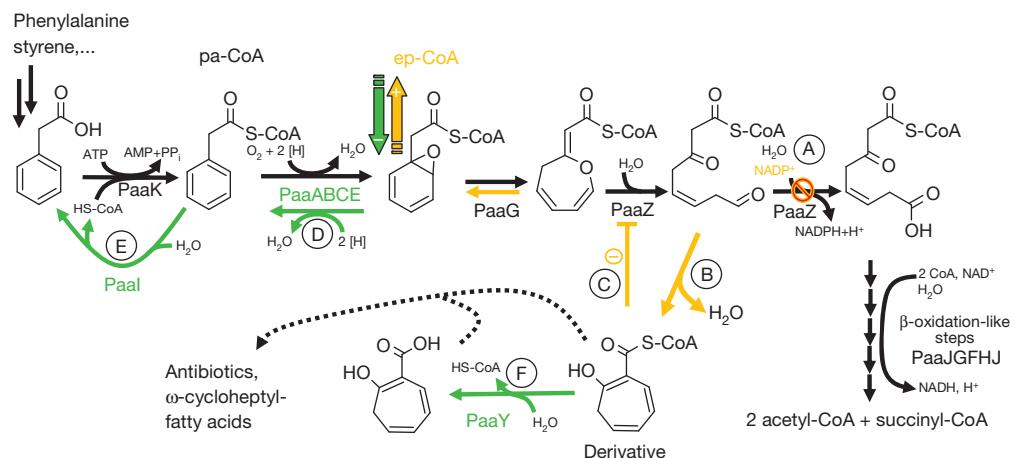


Figure 1 | Overview of reactions and proposed salvage reactions of phenylacetate degradation. Black arrows show catabolic steps leading to acetyl-CoA and succinyl-CoA. When oxidation of the labile aldehyde is impaired (A), the derivative 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA is spontaneously formed (B) and inhibits hydrolytic ring-cleavage (C) leading to ring-1,2-epoxyphenylacetyl-CoA (ep-CoA) accumulation (orange arrows). PaaABCE then prevents overflow of toxic epoxides through deoxygenation (D), assisted by pa-CoA thioesterase PaaI (E), whereas 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA thioesterase PaaY diminishes inhibition (F) (green arrows). Some organisms use the (perhaps CoA-free) derivatives as starting material for antibiotics and fatty acids (dashed arrows). For other enzymes and intermediates, see Supplementary Fig. 1.

arrows). PaaABCE then prevents overflow of toxic epoxides through deoxygenation (D), assisted by pa-CoA thioesterase PaaI (E), whereas 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA thioesterase PaaY diminishes inhibition (F) (green arrows). Some organisms use the (perhaps CoA-free) derivatives as starting material for antibiotics and fatty acids (dashed arrows). For other enzymes and intermediates, see Supplementary Fig. 1.

¹Mikrobiologie, Fakultät Biologie, Universität Freiburg, Schänzelstrasse 1, D-79104 Freiburg, Germany. ²Organische Chemie und Biochemie, Fakultät Chemie, Universität Freiburg, Albertstrasse 21, D-79104 Freiburg, Germany.

One mol PaaE contained 1.9 ± 0.2 mol acid-labile sulphur and 0.73 mol FAD, confirming the presence of one FAD and one [2Fe-2S] binding domain per subunit E. Ultraviolet-visible spectra were consistent with the presence of these cofactors (Supplementary Fig. 4).

To test PaaABCE activity, NAD(P)H and oxygen consumption were monitored in photometric tests and with an oxygen optode (Methods), respectively. Epoxidase activity amounted to ~ 1.0 U mg $^{-1}$ with NADPH and ~ 0.05 U mg $^{-1}$ with NADH (at 30°C and optimal pH 8). As determined by high performance liquid chromatography (HPLC), pa-CoA was consumed ~ 1.5 times faster when ring-1,2-epoxyphenylacetyl-CoA (ep-CoA) was removed through subsequent enzymes PaaG and PaaZ (Fig. 1). PaaABCE exhibited high affinities for NADPH (apparent Michaelis constant $K_m = 23 \pm 3$ μM), pa-CoA ($K_m = 6 \pm 1$ μM) and O_2 ($K_m = 3 \pm 1$ μM). In the absence of pa-CoA, PaaABCE uncoupled and slowly reduced O_2 to water rather than H_2O_2 , with a NADPH oxidation rate of 0.07 ± 0.02 U mg $^{-1}$ (black dashed line, Fig. 2a). Addition of pa-CoA then switched the enzyme from oxidase to oxygenase (Fig. 2a, Supplementary Table 1). The non-physiological substrate benzoyl-CoA was not transformed to detectable products, but apparently first uncoupled and then permanently inactivated PaaABCE (Supplementary Fig. 6). An overview of the main features of PaaABCE is given in Supplementary Table 1.

Surprisingly, when pa-CoA was virtually completely converted to ep-CoA, the NADPH consumption remained significantly higher (dashed purple line, Fig. 2a) than in the initially observed uncoupled state (that is, before pa-CoA addition). When oxygen concentration was monitored, the rate of O_2 consumption was half that of NADPH consumption (Supplementary Table 2, Supplementary Fig. 7). Hence, the produced ep-CoA probably induced NADPH and O_2 consumption. To better understand this behaviour, we took sequential samples from a photometric test and immediately quenched the enzymatic reaction, followed by HPLC and mass spectrometry product analysis. We observed nearly complete back conversion of ep-CoA to pa-CoA when O_2 became limiting. Moreover, pa-CoA was again transformed to ep-CoA after adding O_2 (sample 5 and 6, Fig. 2), indicating that PaaABCE also functions as a deoxygenase (epoxide reductase), whereas spontaneous ep-CoA decay produced 2-hydroxyphenylacetate and CoA (Fig. 2b and Supplementary Fig. 8). Importantly, PaaABCE strictly required NADPH to achieve deoxygenation of ep-CoA (Supplementary Fig. 9). The continuing NADPH consumption in the photometric test thus implied an equilibrium between ep-CoA deoxygenation and pa-CoA reoxygenation that was determined by the absence or presence of oxygen. To further test the deoxygenation hypothesis, we added ^{18}O -labelled ep-CoA to PaaABCE in the presence of NADPH and air-saturated water. We observed the exchange of ^{18}O for ^{16}O in the epoxide over time (Supplementary Fig. 10). On the basis of our experiments and of established knowledge from other extensively studied di-iron enzymes (benzoyl-CoA epoxidase 16 , soluble methane monooxygenase (sMMO) 17), we tentatively propose the following reactions (for steps and intermediates see Fig. 3).

First, we consider PaaABCE in the presence of O_2 , NADPH and pa-CoA (continuous red line, Fig. 2a). On binding of the substrate, the di-iron centre ground state $\text{Fe}^{\text{III}}(\mu\text{-O})\text{Fe}^{\text{III}}$ of PaaA (compound 1, equivalent to compound T of sMMO 18) is NADPH-dependently reduced to $\text{Fe}^{\text{II}}\text{Fe}^{\text{II}}$ (compound 2, equivalent to compound H $_{\text{red}}$ of sMMO 19) with release of water (step I). This state then directly reacts with O_2 and generates a peroxy $\text{Fe}^{\text{IV}}\text{-O-O-Fe}^{\text{IV}}$ species, which rearranges to the key oxygenating intermediate $\text{Fe}^{\text{IV}}(\mu\text{-O})_2\text{Fe}^{\text{IV}}$ ('diamond core') or alternatively $\text{O} = \text{Fe}^{\text{IV}}(\mu\text{-O})\text{Fe}^{\text{IV}}$ in the presence of protons (compound 3, equivalent to compound Q of sMMO $^{20-22}$) (step IIa). This allows for the epoxidation of pa-CoA and regeneration of the di-iron centre to compound 1 (step III), which prevalently occurs in the presence of pa-CoA and oxygen.

Second, PaaABCE in the presence of O_2 , NADPH and ep-CoA (dashed purple line, Fig. 2a). Remarkably, the di-iron centre also deoxygenates ep-CoA, which occurs when pa-CoA is largely used up. Once

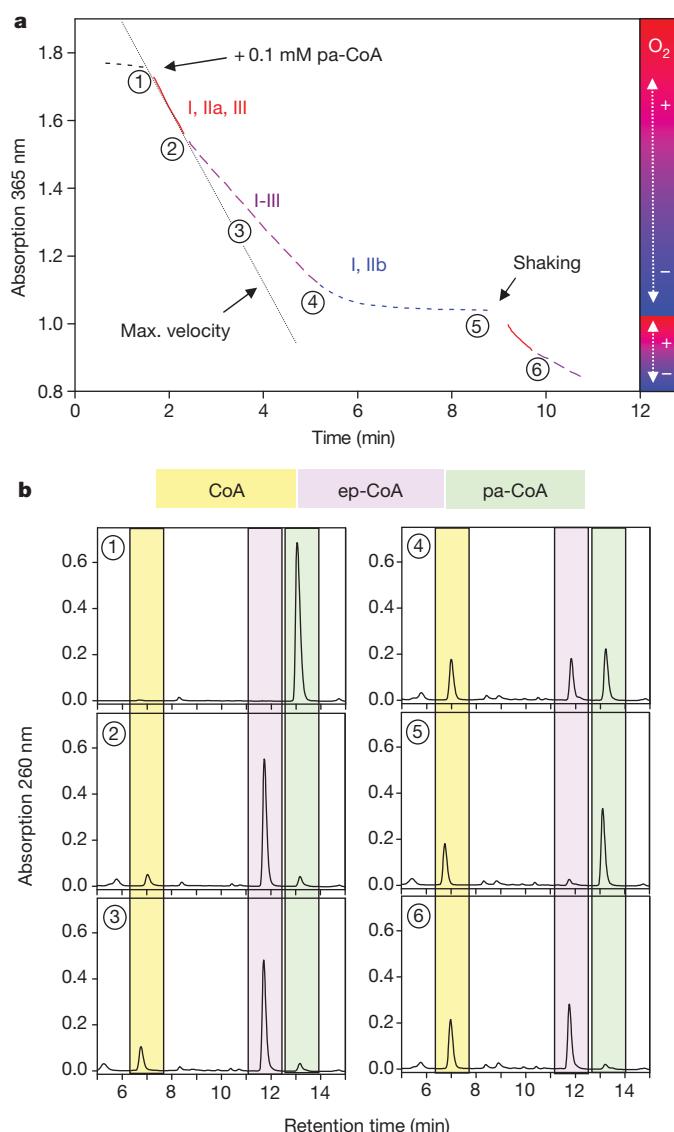


Figure 2 | PaaABCE is a bifunctional monooxygenase/deoxygenase. a, NADPH consumption by PaaABCE (0.25 mg protein added per ml) was monitored photometrically in a 0.5-cm cuvette; maximum velocity (max. velocity indicated by thin line) was attained after addition of pa-CoA (+ 0.1 mM phenylacetyl-CoA). NADPH consumption slowed and finally stopped completely when oxygen was depleted. NADPH consumption was restarted once the cuvette was shaken to introduce oxygen again. A change in absorbance at 365 nm wavelength ($\Delta A_{365\text{nm}}$) of 0.175 is equivalent to consumption of 0.1 mM NADPH ($\varepsilon_{365\text{nm}}$ of $3,500 \text{ M}^{-1} \text{ cm}^{-1}$). Samples 1 to 6 were taken at different time points; their analysis by HPLC is shown in b. Curve sections are illustrated by differently coloured dotted or continuous lines. The correspondingly coloured roman numerals indicate the reactions taking place (shown in Fig. 3). b, RP-HPLC analysis of samples from a. Product peaks were analysed by mass spectrometry. CoA (observed) and 2-hydroxyphenylacetate (not visible, see Supplementary Fig. 8 for proof) are released during spontaneous decay during which no other CoA-esters were observed. Compounds are highlighted with shadings in yellow (CoA), purple (ep-CoA), and green (pa-CoA).

ep-CoA has bound, compound 1 is NADPH-dependently reduced to compound 2 with release of water as described above. Crucially, instead of reacting with O_2 , the reduced di-iron centre may abstract the epoxy-oxygen from ep-CoA with generation of pa-CoA and resting state compound 1 (step IIb). Any formed pa-CoA will be converted back to ep-CoA by steps I, IIa and III. Hence, 2 NADPH are net consumed per 1 O_2 with formation of 2 H_2O , whereas pa-CoA and ep-CoA equilibrate. As observed in our HPLC assays, the reaction rates

favoured the ep-CoA side under oxic conditions, implying somewhat slower (40%, see below) deoxygenation compared to oxygenation (100%) (Supplementary Table 1; sample 3, Fig. 2b). This conclusion is consistent with distinct decreases in NADPH and O₂ consumption rates when pa-CoA was largely used up (Fig. 2a, Supplementary Fig. 7, Supplementary Table 2).

Third, PaaABCE in the absence of O₂ and in the presence of ep-CoA and NADPH (dashed blue line, Fig. 2a). When O₂ is depleted, only NADPH-driven deoxygenation occurs (steps I and IIb), eventually resulting in complete reduction of ep-CoA-derived oxygen to water via step I. Hence, 1 ep-CoA is converted to 1 pa-CoA under NADPH-consuming formation of 1 H₂O until the ep-CoA pool is depleted (sample 4 and 5, Fig. 2b). In total, 1 NADPH is oxidized for each initially produced ep-CoA.

According to our model, PaaABCE ultimately reduces all oxygen to water via step I (Fig. 3). Hence, 2 NADPH should be consumed for each O₂ reduced to 2 H₂O, independent of the amount of pa-CoA added. To test this counterintuitive hypothesis, we started oxygen-limited (concentration of O₂ = 0.235 mM) photometric assays by adding 0.05, 0.1, 0.2 and 0.4 mM pa-CoA to the assay mixture. Indeed, a constant stoichiometry of ~1.8 NADPH oxidized: 1 O₂ added resulted. The small deviation from the expected stoichiometry of 2:1 can be explained by spontaneous partial decay of ep-CoA sparing 1 NADPH for each decomposed ep-CoA. In contrast, when produced ep-CoA was enzymatically removed, only 1 NADPH was consumed per 1 pa-CoA added (Supplementary Table 3). Photometric assays were analysed based on our reaction model, and a specific activity of ~0.4 U mg⁻¹ was calculated for ep-CoA deoxygenation (~40% compared to pa-CoA oxygenation). The apparent K_m value of PaaABCE for ep-CoA was 17 ± 2 μM.

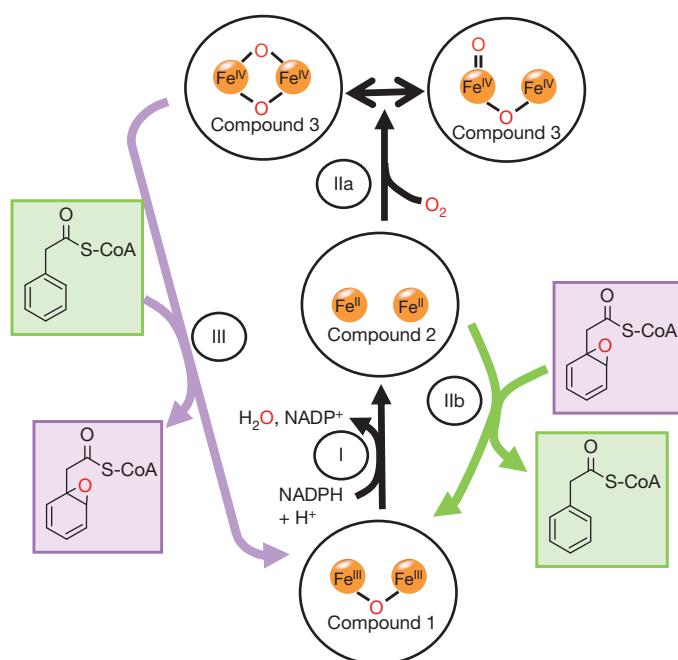


Figure 3 | Simplified scheme of proposed reactions catalysed by PaaABCE and states of the catalytic di-iron centre. Di-iron compounds are proposed in analogy to sMMO (see text). I, NADPH-dependent reduction of the ground state di-iron core to diferric compound 2 under formation of water. IIa, the reduced state directly reacts with molecular oxygen and produces a high-valent intermediate (compound 3). IIb, Alternatively, the reduced state may abstract the epoxy-oxygen to produce pa-CoA and compound 1. III, compound 3 *ortho*-oxygenates pa-CoA to produce ep-CoA and compound 1. Note that high-valent compound 3 of sMMO has not yet been unambiguously identified (indicated with double-headed arrow). Green arrows indicate pa-CoA formation, purple arrows indicate ep-CoA formation. Pa-CoA and ep-CoA are shown in green and purple boxes, respectively.

At first glance, deoxygenation of the epoxide may be incidental and counterproductive; still, this bifunctionality may also imply biological aims, as suggested by the high affinity for ep-CoA. But why must the cells prevent a rise in concentration of ep-CoA? First, the labile epoxide rapidly decays to unproductive 2-hydroxyphenylacetate (Supplementary Fig. 8) that cannot be used³. Second, epoxides are notoriously reactive towards DNA and protein²³ and severely hamper cell growth²⁴. Normally, hydrolytic ring-cleavage by PaaZ ultimately removes ep-CoA. However, we found that a product (2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA, see Fig. 1) that is rapidly formed when PaaZ function is impaired by NADP⁺ shortage⁶ is a potent ring-cleavage inhibitor (Supplementary Fig. 11). Note that PaaZ specifically requires NADP⁺ (apparent K_m of 17 ± 3 μM). Activity with NAD⁺ was <5%; previous reports⁶ showing activity with NAD⁺ were based on absorption increase due to product rather than NADH formation (see also discussion in Supplementary Information). In *Escherichia coli*, the cellular NADP⁺ concentration was reported to be 2 μM, in contrast to 120 μM NADPH²⁵, and cells maintain a high NADPH/NADP⁺ ratio independent of the growth substrate used. As bacteria inevitably encounter situations in which the NADPH formation exceeds its demand²⁶, short-term lacks of NADP⁺ may well trigger ep-CoA accumulation (Fig. 1). We hypothesize that in this case, pa-CoA and ep-CoA both compete for PaaABCE, which catalyses an ongoing NADP⁺-restoring oxygenation/deoxygenation cycle that prevents indefinite epoxide rise. It is tempting to speculate that certain pathogens like *Burkholderia cenocepacia* found alternatives for epoxide disposal, for example, as part of the pathogen's arsenal against host cells, which may explain increasing evidence for PaaABCE as a virulence factor^{10,11}.

The detoxification process, however, proceeds more efficiently if the concentration of the competing substrate pa-CoA does not rise in consequence of ep-CoA deoxygenation. We speculate that thioesterase PaaI, which exhibits high affinity for pa-CoA (apparent K_m ≈ 10 μM)²⁷, removes excess pa-CoA. To test this, we compared relevant pathway intermediates as substrates for PaaI (among others, see Supplementary Table 4) and found an atypical narrow substrate spectrum with reasonable activity exclusively for pa-CoA (1.6 ± 0.1 U mg⁻¹; 22 °C). PaaI may thus act as a pa-CoA thioesterase that supports PaaABCE in controlling ep-CoA concentration (Fig. 1). Aptly, we found that PaaY constitutes a second, previously undescribed type of thioesterase, which specifically hydrolyses inhibitory 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA (7.6 ± 0.6 U mg⁻¹, apparent K_m 35 ± 15 μM; 22 °C) (Fig. 1, see Supplementary Table 4 for substrate specificity) to recover CoA and counteract ring-cleavage inhibition (Supplementary Fig. 11). Hence, PaaY may relieve epoxide accumulation and have a stimulatory effect on intracellular phenylacetate consumption. As expected, both thioesterases seem induced under growth on phenylacetate (Supplementary Table 4). We note that some organisms found alternatives for 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA disposal and may use it as a starting material for the synthesis of ω -cycloheptyl fatty acids and distinct antibiotics⁶ (Supplementary Fig. 1).

Although the versatile di-iron proteins catalyse diverse reactions with complex underlying chemistry, such as ribonucleotide reduction, methane monooxygenation or Δ^9 -desaturation of stearoyl-acyl carrier protein¹⁴, they are commonly monofunctional, and none is able to recycle its substrate by using the primary product in an alternative reaction. Therefore, PaaABCE is outstanding among the di-iron proteins. The deoxygenation reaction is, from a thermodynamical perspective, enabled by the reactive epoxide that is significantly higher in energy than, for example, the phenolic alcohol that would arise from ring-hydroxylation. Of particular interest for further studies of PaaABCE are the di-iron species responsible for the key oxygenating and deoxygenation reactions. Apart from the still not clearly resolved di-iron compounds of sMMO, there are numerous other possibilities for reactive di-iron species, as reports of artificial biomimetic complexes indicate^{28,29}.

In summary, bifunctional PaaABCE acts as a phenylacetyl-CoA ring-1,2-epoxidase/ring-1,2-epoxyphenylacetyl-CoA deoxygenase that may harbour an effective intrinsic safety mechanism allowing for undelayed restraint of epoxide formation. Thioesterases PaaI and PaaY assist detoxification and prevent depletion of the CoA-pool. Recruitment of two separate thioesterases with an uncommon narrowed substrate spectrum for ring-closed CoA esters might reflect the need to circumvent release of reactive ring-epoxides from ep-CoA. It seems that usage of labile and toxic intermediates in this biological system imposes elaborate escape mechanisms on the cells under stress conditions, allowing for recycling of phenylacetate that can be catabolized again once the distress eases. The intriguing question remains whether in other cases monooxygenases from the di-iron family (for example, alkene epoxidase³⁰, benzoyl-CoA epoxidase¹⁶) similarly exploit the highly energetic epoxides in analogous reactions and cooperate with other enzymes. Deoxygenation may then be a general means to control epoxide levels.

METHODS SUMMARY

Recombinant maltose-binding protein (mbp)-tagged Paa_{mbp}ABCE from *Pseudomonas* sp. strain Y2 and Paa_{mbp}I and Paa_{mbp}Y from *E. coli* K12 were heterologously expressed in *E. coli* and purified by amylose resin chromatography. Pa-CoA was chemically synthesized¹; all other pathway intermediates were enzymatically synthesized and purified by reversed-phase (RP)-HPLC.

For detection of metals, recombinant PaaABCE was analysed using inductively coupled plasma emission spectroscopy. PaaABCE was denatured to isolate the FAD cofactor which was identified by RP-HPLC. The acid-labile sulphur content was determined by photometrical quantification of methylene blue formed through reaction of released sulphide from denatured protein and N,N'-dimethyl-*p*-phenylenediamine, using Na₂S·9H₂O as the standard. Enzyme composition was determined by quantifying the spot intensities for each subunit on SDS gels. Enzymatic activities and apparent *K*_m values for PaaABCE were determined at 30 °C by following NADPH consumption in photometric assays at 365 nm or oxygen consumption using an oxygen optode. PaaI and PaaY activities and apparent *K*_m values were determined at 22 °C using RP-HPLC assays with optical detection of CoA and CoA-esters that were separated with a C18-E column using an ammonium acetate-buffered (pH 6.8) acetonitrile gradient.

For mass spectrometry (MS) measurements, RP-HPLC samples were directly injected into the TurbolonSpray ion source of an API 2000 LC/MS/MS mass spectrometer. EPR spectra of PaaABCE were recorded with a Bruker EMX 1/6 spectrometer operating at X-band (9.4 GHz).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 July 2011; accepted 16 January 2012.

Published online 7 March 2012.

- Teufel, R. *et al.* Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc. Natl Acad. Sci. USA* **107**, 14390–14395 (2010).
- Luengo, J. M., Garcia, J. L. & Olivera, E. R. The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications. *Mol. Microbiol.* **39**, 1434–1442 (2001).
- Ismail, W. *et al.* Functional genomics by NMR spectroscopy. Phenylacetate catabolism in *Escherichia coli*. *Eur. J. Biochem.* **270**, 3047–3054 (2003).
- Grishin, A. M. *et al.* Structural and functional studies of the *Escherichia coli* phenylacetyl-CoA monooxygenase complex. *J. Biol. Chem.* **286**, 10735–10743 (2011).
- Fernández, C., Ferrández, A., Minambres, B., Diaz, E. & García, J. L. Genetic characterization of the phenylacetyl-coenzyme A oxygenase from the aerobic phenylacetic acid degradation pathway of *Escherichia coli*. *Appl. Environ. Microbiol.* **72**, 7422–7426 (2006).
- Teufel, R. *et al.* Studies on the mechanism of ring hydrolysis in phenylacetate degradation: a metabolic branching point. *J. Biol. Chem.* **286**, 11021–11034 (2011).
- Geng, H. & Belas, R. Expression of tropodithietic acid biosynthesis is controlled by a novel autoinducer. *J. Bacteriol.* **192**, 4377–4387 (2010).
- Thiel, V. *et al.* Identification and biosynthesis of tropone derivatives and sulfur volatiles produced by bacteria of the marine Roseobacter clade. *Org. Biomol. Chem.* **8**, 234–246 (2010).

- Moore, B. S. *et al.* Biosynthetic studies of omega-cycloheptyl fatty acids in *Alicyclobacillus cycloheptanicus*. Formation of cycloheptanecarboxylic acid from phenylacetic acid. *J. Org. Chem.* **62**, 2173–2185 (1997).
- Law, R. J. *et al.* A functional phenylacetic acid catabolic pathway is required for full pathogenicity of *Burkholderia cenocepacia* in the *Caenorhabditis elegans* host model. *J. Bacteriol.* **190**, 7209–7218 (2008).
- Hamlin, J. N., Bloodworth, R. A. & Cardona, S. T. Regulation of phenylacetic acid degradation genes of *Burkholderia cenocepacia* K56-2. *BMC Microbiol.* **9**, 222 (2009).
- Khomyakova, M., Bükmel, O., Thomas, L. K., Erb, T. J. & Berg, I. A. A methylaspartate cycle in halarchaea. *Science* **331**, 334–337 (2011).
- Leahy, J. G., Batchelor, P. J. & Morcomb, S. M. Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* **27**, 449–479 (2003).
- Nordlund, P. & Eklund, H. Di-iron-carboxylate proteins. *Curr. Opin. Struct. Biol.* **5**, 758–766 (1995).
- Rather, L. J., Knapp, B., Haehnel, W. & Fuchs, G. Coenzyme A-dependent aerobic metabolism of benzoate via epoxide formation. *J. Biol. Chem.* **285**, 20615–20624 (2010).
- Rather, L. J. *et al.* Structure and mechanism of the diiron benzoyl-coenzyme A epoxidase BoxB. *J. Biol. Chem.* **286**, 29241–29248 (2011).
- Rosenzweig, A. C., Frederick, C. A., Lippard, S. J. & Nordlund, P. Crystal structure of a bacterial non-haem iron hydroxylase that catalyses the biological oxidation of methane. *Nature* **366**, 537–543 (1993).
- Lee, S. K., Nesheim, J. C. & Lipscomb, J. D. Transient intermediates of the methane monooxygenase catalytic cycle. *J. Biol. Chem.* **268**, 21569–21577 (1993).
- Rosenzweig, A. C., Nordlund, P., Takahara, P. M., Frederick, C. A. & Lippard, S. J. Geometry of the soluble methane monooxygenase catalytic diiron center in two oxidation states. *Chem. Biol.* **2**, 409–418 (1995).
- Shu, L. *et al.* An Fe₂^{IV}O₂ diamond core structure for the key intermediate Q of methane monooxygenase. *Science* **275**, 515–518 (1997).
- Rowe, G. T., Rybak-Akimova, E. V. & Caradonna, J. P. Unraveling the reactive species of a functional non-heme iron monooxygenase model using stopped-flow UV-vis spectroscopy. *Inorg. Chem.* **46**, 10594–10606 (2007).
- Rowe, G. T., Rybak-Akimova, E. V. & Caradonna, J. P. Heterolytic cleavage of peroxide by a diiron compound generates metal-based intermediates identical to those observed with reactions utilizing oxygen-atom-donor molecules. *Chemistry* **14**, 8303–8311 (2008).
- Phillips, D. H. & Farmer, P. B. Evidence for DNA and protein binding by styrene and styrene oxide. *Crit. Rev. Toxicol.* **24**, S35–S46 (1994).
- Park, J. B. *et al.* The efficiency of recombinant *Escherichia coli* as biocatalyst for stereospecific epoxidation. *Biotechnol. Bioeng.* **95**, 501–512 (2006).
- Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chem. Biol.* **5**, 593–599 (2009).
- Fuhrer, T. & Sauer, U. Different biochemical mechanisms ensure network-wide balancing of reducing equivalents in microbial metabolism. *J. Bacteriol.* **191**, 2112–2121 (2009).
- Song, F. *et al.* Structure, function, and mechanism of the phenylacetate pathway hot dog-fold thioesterase Paal. *J. Biol. Chem.* **281**, 11028–11038 (2006).
- Johansson, A. J., Noack, H., Siegbahn, P. E., Xue, G. & Que, L. Jr. Observed enhancement of the reactivity of a biomimetic diiron complex by the addition of water — mechanistic insights from theoretical modeling. *Dalton Trans.* 6741–6750 (2009).
- Xue, G., De Hont, R., Munck, E. & Que, L. Jr. Million-fold activation of the [Fe₂(μ-O₂)₂] diamond core for C–H bond cleavage. *Nature Chem.* **2**, 400–405 (2010).
- Gallagher, S. C., Cammack, R. & Dalton, H. Alkene monooxygenase from *Nocardia corallina* B-276 is a member of the class of dinuclear iron proteins capable of stereospecific epoxidation reactions. *Eur. J. Biochem.* **247**, 635–641 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Deutsche Forschungsgemeinschaft for support; V. Mascaraque, G. Heijman and W. Ismail for their contributions in the initial stage of this work; M. Voss, M. Weiß, I. Sachelaru and T. Spatzl for technical assistance; M. Müller and V. Brecht for mass spectrometry analysis; and I. Berg and H. Heider for critical comments.

Author Contributions R.T. and T.F. performed research; R.T., T.F. and G.F. designed research and analysed data; and R.T. and G.F. wrote the paper.

Author Information For *E. coli* K12, the NCBI database accession numbers are as follows: PaaY, NP_415918.1; Paal, NP_415914.1. For *Pseudomonas* sp. strain Y2, the EMBL database numbers are as follows: PaaA, CAD76929.1; PaaB, CAD76932.1; PaaC, CAD76933.1; PaaD, CAD76935.1; PaaE, CAD76937.1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.F. (georg.fuchs@biologie.uni-freiburg.de).

METHODS

Materials. ^{14}C -phenylacetate was obtained from Biotrend Chemikalien GmbH. All other materials were obtained from previously described sources⁶.

Syntheses. Phenylacetyl-CoA was chemically synthesized as described before¹. Pathway CoA-ester intermediates were enzymatically synthesized and purified in large scale reactions. For production of the derivative 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA, a reaction mixture (4 ml) containing 0.06 mg PaaZ-E256Q (variant enzyme lacking aldehyde dehydrogenase activity), 0.3 mg PaaABCE, 0.03 mg PaaG, 1.5 mM NADPH, 50 mM Tris-HCl (pH 8.0) and 0.6 mM pa-CoA was stirred for 7 min at 30 °C. For other pathway intermediates, similar assays were performed using PaaABCE, PaaG, PaaZ and PaaJ as required. For ^{16}O -ep-CoA, a reaction mixture (10 ml) containing 2.3 mg PaaABCE, 1.5 mM NADPH, 50 mM Tris-HCl (pH 8.0) and 0.7 mM pa-CoA was stirred for 6 min at 22 °C. For ^{18}O -ep-CoA, a reaction tube (volume 7.5 ml) containing a 3 ml reaction mixture of 0.3 mM pa-CoA, 1 mM NADPH and 50 mM Tris-HCl (pH 8.0) was closed and degassed. A total volume of 6 ml $^{18}\text{O}_2$ gas was subsequently injected and the assay stirred for 2 min. The reaction was started by addition of 0.05 mg degassed PaaABCE and stirred for additional 3 min at 22 °C. After incubation, all reactions were stopped, the CoA-esters were purified as described before¹ and kept frozen (-70°C). The CoA-esters were dissolved in water or 20 mM 2-(N-morpholino)ethanesulphonic acid (MES)-KOH pH 6.0 (ep-CoA, oxepin-CoA) before use. Other CoA esters were chemically synthesized or purchased as described before⁶. CoA thioesters were quantified as previously described⁶.

RP-HPLC analysis and CoA-ester and radioactivity detection. Detection and separation of CoA-esters in different samples were carried out by RP-HPLC as described before¹. The products of the conversion of [^{14}C]phenylacetate were analysed by flowthrough solid state scintillation counting of ^{14}C , as described previously³¹. The radioactive samples were separated using a Luna 5U C18 (2) 100A column (250 × 4.60 mm, 5 μm) (Phenomenex). The column was developed at a flow rate of 1 ml min⁻¹ by a gradient from 2% acetonitrile in 40 mM phosphate buffer (pH 4.0) to 70% acetonitrile within 32 min. Buffer was gradually exchanged for water during the gradient.

Molecular biological techniques. All *in silico* cloning steps were performed with the program Clone Manager 7.11 (Scientific & Educational Software). The cloning procedures for *paaABCDE*, *paaD*, *paaJ* and *paaG* from *Pseudomonas* sp. strain Y2, and *paaZ* and mutated *paaZ* (encoding PaaZ-E256Q) from *E. coli* K12, and *paaK* from *Thermus thermophilus* were described before^{1,6,32}. Note that PaaD separated during purification of PaaABCDE. Chromosomal DNA was extracted using standard techniques. Genes for *paaY* from *E. coli* K12 were amplified using forward primer 5'-GGCGAATTCCCAATTATCATAGATAGACG-3' and reverse primer 5'-GCTAAGCTTAACTCTGCTTATAATTAAAGG-3'. Genes for *paaI* from *E. coli* K12 were amplified using forward primer 5'-CTGAAGAGCTCTATGA GTCATAAGGCCTGGC-3' and reverse primer 5'-CCGTCTAGATAAAAGGCT TCACGCATCAG-3'. Introduced restriction sites are underlined; the forward primers contained EcoRI (*paaY*) or XmnI (*paaI*) restriction sites, the reverse primers HindIII (*paaY*) or XbaI (*paaI*) restriction sites. PCR was performed with a mixture (5:1) of Taq polymerase and Pfu polymerase (Genaxxon) for 30 cycles, including denaturation for 45 s at 95 °C, annealing for 45 s at 65 °C (*paaI*) or at 55 °C (*paaY*), polymerization for 1 min at 72 °C and a final extension at 72 °C for 5 min. The PCR products were isolated and cloned into the expression vector pMAL-c2X (New England Biolabs), which adds a maltose-binding protein (mbp) to the amino terminus of the expressed proteins.

Analysis by mass spectrometry (MS). Samples of RP-HPLC fractions were directly injected into the TurboIonSpray ion source of an API 2000 LC/MS/MS mass spectrometer (Applied Biosystems). For online mass detection only the Q1 quadrupole was used. Measurements covered the range of *m/z* from 800 to 1,000.

Bacterial strains, growth conditions, and protein purification. PaaABCE, PaaG, PaaK, PaaZ and PaaZ-E256Q were produced, purified, and stored as described before^{1,6,32}. The production, purification and storage of the heterologously produced proteins PaaI and PaaY was performed as described for PaaABCE¹.

Protein analysing methods. Enzyme fractions were analysed by SDS-12.5% polyacrylamide gel electrophoresis³³. Proteins were visualized by Coomassie brilliant blue R-250 staining. Protein concentrations were determined by the Bradford method³⁴ with bovine serum albumin as a standard. Bands of PaaABCE subunits were excised, and the peptides digested and analysed by LC-MS/MS as described before³⁵. The relative protein amounts of PaaA, B, C and E of different enzyme preparations were calculated by determining the spot intensities of each subunit on a SDS-gel using a ChemiDoc XRS Universal Hood II and Quantity one software (Version 4.6.3) (both Biorad). To determine the stoichiometric composition of the purified complex, the spot intensity was correlated with the molecular mass of each protein (subunit A was corrected for the size of untagged protein). For iron and cofactor analysis, the absolute amounts of PaaA and PaaE within the respective sample were then calculated by means of Bradford

protein quantification and correlated with the determined amount of iron, acid-labile sulphur, and FAD.

Enzyme measurements. Enzymatic reactions were carried out at 30 °C (PaaABCE, PaaZ) or 22 °C (PaaI, PaaY) and were buffered with 50 mM Tris-HCl (pH 8.0) if not stated otherwise. At least two independent measurements were performed. The reactions were stopped with 1% formic acid resulting in a pH drop to pH 4.5–5 (stronger acidification rapidly destroyed ep-CoA), and denatured protein was subsequently removed by centrifugation. For PaaI and PaaY, the activities were corrected for the size of untagged protein. The apparent *K_m* value of PaaZ for NADP⁺ was determined in photometric tests using variable amounts of NADP⁺ (0.025–1 mM) and constant amounts of oxepin-CoA (0.2 mM) in the presence of excess PaaY.

A spectrophotometric assay was generally used for the measurement of PaaABCE activity, in which the oxidation of NADPH (or NADH for cofactor specificity) was followed spectrophotometrically at 365 nm ($\epsilon_{365\text{nm}}$ for NADPH of 3,500 M⁻¹ cm⁻¹ and for NADH of 3,400 M⁻¹ cm⁻¹) in 1.0-cm or 0.5-cm cuvettes. Occasionally, separately produced PaaD was added to test for possible involvement in catalysis. To test benzoyl-CoA as a non-physiological substrate for PaaABCE (Supplementary Fig. 6), NADPH oxidation was monitored in a photometric test (300 μl) containing 0.12 mg PaaABCE and 0.5 mM NADPH. The reaction was started with 0.05–0.3 mM benzoyl-CoA. After NADPH oxidation nearly halted, 0.2 mM pa-CoA was added, followed by subsequent addition of fresh PaaABCE (0.1 mg ml⁻¹). NADPH consumption by PaaABCE was also measured photometrically in the absence of substrate. For this purpose, a reaction mixture (0.2 ml) containing 1 mM NADPH and 0.4 mg PaaABCE was monitored at 30 °C. The cuvette was covered with paraffin oil to prevent oxygen diffusion and the assay recorded until NADPH consumption stopped due to lack of oxygen. Approximately 2 NADPH were consumed per O₂. Also, addition of 0.1 mg catalase had no effect on NADPH consumption, indicating that O₂ was reduced to water rather than to H₂O₂.

The NADPH consumption was determined for varying pa-CoA concentrations of 0.05, 0.1, 0.2 and 0.4 mM in air-saturated water at 30 °C (0.235 mM dissolved O₂) (Supplementary Table 3). For this purpose, photometric tests (0.8 ml) containing 0.2 mg PaaABCE and 1 mM NADPH were monitored until NADPH oxidation stopped. The cuvettes were covered with paraffin oil. Decomposed ep-CoA would cause somewhat lower oxidation of NADPH due to trapped oxygen in 2-hydroxyphenylacetate according to the reaction model in Fig. 3. Hence, HPLC samples were withdrawn and the amount of decomposed ep-CoA was calculated by quantifying the peak area of released CoA. The values were then used to correct the expected amounts of oxidized NADPH. Furthermore, to determine the pa-CoA:NADPH stoichiometry when ep-CoA is removed through subsequent steps, PaaG and PaaZ-E256Q mutant enzyme lacking aldehyde dehydrogenase function (instead of PaaZ to avoid recycling of NADPH) were added to photometric tests. The presence of PaaY assured that arising 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA (which absorbs at 365 nm) was removed.

To determine the pH optimum of PaaABCE, reaction mixtures at 22 °C were buffered either with 100 mM MES-KOH (for pH values 6.5–7.0) or 100 mM Tris-HCl (for pH values 7.5–9.0) and were started by addition of 0.1 mM pa-CoA. NADPH oxidation was followed spectrophotometrically and the specific activities were compared between pH 6.5 to 9 in 0.5 pH steps.

The apparent *K_m* values for pa-CoA, ep-CoA and NADPH were determined in photometric tests. For pa-CoA, variable amounts of pa-CoA (5–100 μM) were added to a reaction mixture containing saturating NADPH (130 μM) and O₂ (0.235 mM). The *K_m* value for NADPH was determined by adding variable amounts of NADPH (10–250 μM) to the assay, while the pa-CoA concentration was kept constant (100 μM). The *K_m* value for ep-CoA was indirectly determined by adding variable amounts of pa-CoA (2.5–100 μM) to assays with constant NADPH concentrations (0.5 mM). The NADPH decrease eventually slowed down and adopted a second constant rate when pa-CoA was largely converted to ep-CoA. The velocity of this second rate was analysed for different amounts of produced ep-CoA, and the apparent *K_m* value was calculated. The apparent *K_m* value for O₂ was determined by observing the oxygen consumption in an enzymatic test (0.5 ml), using a MICROX TX3 fibre-optic oxygen meter ('oxygen optode') (PreSens GmbH), which was calibrated at 30 °C with air-saturated water (100% corresponding to 0.235 mM O₂)¹⁵ and freshly prepared 60 mM dithionite (0%). The reaction mixture contained 0.28 mg PaaABCE, 0.03 mg PaaG, 0.1 mg PaaZ, 0.7 mM pa-CoA, 1 mM NADPH and 1 mM NADP⁺. The apparent *K_m* value was graphically determined and corresponded to the curve point at which half-maximal velocity was attained.

To test the effect of ep-CoA withdrawal on PaaABCE catalysed substrate (pa-CoA) consumption, subsequent enzymes PaaG and PaaZ that ultimately remove ep-CoA were added to the enzymatic assay. The assays contained 1 mM NADPH,

1 mM NADP⁺, 0.5 mM pa-CoA, and either only 0.03 mg PaaABCE or 0.03 mg PaaABCE plus 0.05 mg PaaG, and 0.3 mg PaaZ.

To compare the velocities of oxygen and NADPH consumption (Supplementary Table 2, Fig. 2, Supplementary Fig. 7), enzymatic assays (800 µl) containing 1 mM NADPH and 0.12 mg PaaABCE were started by addition of 0.1 mM pa-CoA at 30 °C and were either monitored for NADPH decrease (photometrically) or for oxygen consumption (using a calibrated oxygen optode, see above).

Also, oxygen consumption by PaaABCE was monitored under pa-CoA/NADPH limitation with following RP-HPLC product analysis at different time points (Supplementary Fig. 9). The assay (1.3 ml) at 30 °C contained 0.23 mg PaaABCE, 0.22 mM NADPH, and was started by addition of 0.1 mM pa-CoA. The reaction tube was covered with paraffin oil to prevent oxygen diffusion. After oxygen consumption stopped, 0.2 mM NADPH was added. Different samples (1–4) were taken as indicated in Supplementary Fig. 9, immediately stopped, and analysed by RP-HPLC.

Specificity of thioesterases PaaI and PaaY. To determine thioesterase activity of PaaI and PaaY (Supplementary Table 4), formation of CoA from different substrates was quantified by determining the relative peak areas of substrates and products at different time points in RP-HPLC assays at 22 °C. For PaaI activity, assays (0.15 ml) containing 0.08 mg PaaI and 0.25–0.5 mM pa-CoA were incubated for 2 min, stopped, and analysed. For PaaY activity, reaction mixtures (0.1 ml) containing 6 µg PaaY and 0.5 mM 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA were incubated for 2 min, stopped, and analysed. Other substrates (each 0.5 mM) were tested accordingly (see Supplementary Table 4). To determine the apparent *K_m* value of PaaY for 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA, varying amounts (0.02, 0.05, 0.2 and 1 mM) of it were added to enzymatic tests (25–100 µl) containing small amounts of PaaY (0.6–1.2 µg). The reactions were stopped after 2 min and analysed by RP-HPLC. To test for induction of thioesterase activity during growth on phenylacetate, *E. coli* K12 cells were grown on phenylacetate as sole carbon and energy source as described previously³. Similarly, *E. coli* cells were grown on glucose as sole carbon and energy source complemented with vitamins and trace elements in phosphate-buffered mineral salt medium³⁶. Cells were harvested during exponential growth phase and disrupted using a French pressure cell. After ultracentrifugation, the supernatant (cell extract) was used to determine thioesterase activity in HPLC assays (see above) with pa-CoA or 2-hydroxycyclohepta-1,4,6-triene-1-carboxyl-CoA as substrates.

Effect of PaaY on ring-cleavage. To test the effect of PaaY on ring-cleavage (Supplementary Fig. 11), a time course of oxepin-CoA (0.5 mM) consumption and product formation by recombinant purified PaaZ (4.1 µg ml⁻¹) was recorded using RP-HPLC either in the absence of NADPH and PaaY, in the presence of NADP⁺ (1.5 mM) and the absence of PaaY, or in the presence of NADP⁺ (1.5 mM) and PaaY (44 µg ml⁻¹).

Separation of substrates and products of PaaABCE and MS analysis. To observe the enzymatic formation of ep-CoA and pa-CoA during a photometric test (1.5 ml), a reaction mixture containing 0.1 mM pa-CoA, 1 mM NADPH and 0.25 mg ml⁻¹ PaaABCE was monitored at 365 nm. After NADPH oxidation stopped, the cuvette was vigorously shaken to supply oxygen. Different samples were withdrawn (Fig. 2) and the samples analysed by RP-HPLC. CoA-ester peaks were collected for mass spectrometry analysis.

¹⁸O labelling experiments. (Shown in Supplementary Fig. 10.) To determine the PaaABCE-catalysed deoxygenation of ¹⁸O-ep-CoA and subsequent reformation of ¹⁶O-ep-CoA in the absence of pa-CoA, a reaction mixture (620 µl) containing 0.12 mM ¹⁸O-ep-CoA and 1 mM NADPH was started by addition of 0.09 mg PaaABCE and stirred at 22 °C. Samples were taken before reaction start and after 1, 2, 4 and 8 min. As a control, two samples after 2 and 4 min were taken from a second assay without NADPH. Also, an assay (800 µl) containing 0.2 mg PaaABCE and 1 mM NADPH was started by addition of 0.1 mM pa-CoA in 80% H₂¹⁸O/20% H₂¹⁶O. After conversion of pa-CoA to ep-CoA, deoxygenation/reoxygenation was observed for additional 2 min before a sample was taken in which no ¹⁸O-ep-CoA was detected. Ep-CoA was purified by RP-HPLC and frozen in liquid nitrogen until MS measurement.

Detection of 2-hydroxyphenylacetate as decay product of ep-CoA. (Shown in Supplementary Fig. 8.) To show that 2-hydroxyphenylacetate arises from spontaneous ep-CoA decay, phenylacetate-CoA ligase PaaK (1 mg ml⁻¹) was used to enzymatically synthesize ¹²C/¹⁴C-pa-CoA from a mixture of ¹²C/¹⁴C-phenylacetate. A sample was analysed by HPLC using UV and radioactivity detection before NADPH (0.5 mM) and PaaABCE (0.7 mg ml⁻¹) were added. The reaction tube was incubated at 40 °C and vigorously shaken to supply oxygen. Further samples were taken after 1, 3 and 10 min and analysed accordingly.

Inhibition by EDTA. To test the effect of metal chelating agents, 0.17 mg PaaABCE were incubated with 50 mM EDTA (pH 8.0) for 3 h at 30 °C. As a

control, equal protein amounts were incubated without EDTA. Both fractions were then used for photometric activity tests as described above.

Determination of the FAD cofactor. Purified PaaABCE (94 µg) was denatured for 1 h in the dark with 3 µl of 70% HClO₄ to liberate cofactors. The sample was centrifuged (10,000g, 10 min, 4 °C) and 100 µl of the yellowish supernatant containing the isolated cofactor were transferred to 83 µl of 1 M sodium phosphate buffer, pH 7.1 and analysed by RP-HPLC using a C18-E column (LiChrospher 100 RP, end-capped, 5 µm, 125 × 4 mm) and a flow rate of 1 ml min⁻¹. The retention times of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) under isocratic conditions in 9% acetonitrile and 20 mM ammonium acetate (pH 6.0) were 5.5 and 8.5 min, respectively. A distinct peak eluted after 5.5 min retention time. The cofactor was identified as FAD by retention time, and the amount of cofactor extracted was determined by comparison of the peak area with that of FAD standard solutions (0.05–10.0 µg FAD per injection).

Determination of acid-labile sulphur content. The acid-labile sulphur content of (dithioerythritol-free) PaaABCE was determined by denaturing PaaABCE in alkaline medium under release of sulphide, which was coprecipitated with Zn(OH)₂ as ZnS. Acidification then produced H₂S, which was condensed with N,N'-dimethyl-p-phenylenediamine yielding methylene blue, as quantified photometrically at 670 nm, using Na₂S·9H₂O as the standard³⁷.

Spectroscopic analysis of PaaA_{mbp}BCE. (Shown in Supplementary Fig. 4.) Ultraviolet-visible spectra against the blank buffer solution were recorded using a Cary Bio100 spectrophotometer (Varian). A spectrum of oxidized PaaA_{mbp}BCE (0.24 mg ml⁻¹) was monitored in the range 220–600 nm. To visualize FAD cofactors and [2Fe-2S]-clusters, 5.5 mg ml⁻¹ PaaA_{mbp}BCE were recorded in the range 320–600 nm and then completely reduced by addition of 0.7 mM dithionite. Reduced enzyme was then exposed to air to re-oxidize PaaA_{mbp}BCE over time, and further spectra were recorded in 5-min time intervals.

EPR measurements. (Shown in Supplementary Fig. 5.) EPR measurements were conducted with a Bruker EMX 1/6 spectrometer operating at X-band (9.2 GHz). The sample temperature was controlled with an Oxford Instruments ESR-9 helium flow cryostat. The magnetic field was calibrated using a strong pitch standard. The isolated protein was reduced with the addition of a few grains of dithionite.

Metal detection. For detection of metals, 7.1 mg of recombinant purified PaaABCE was analysed by inductively coupled plasma emission spectroscopy (ICP-OES) as described before³⁸.

Computational analysis. (Shown in Supplementary Fig. 2.) Apparent *K_m* and *v_{max}* values were calculated using GraphPad Prism4 software. The BLASTP³⁹ searches were performed via the NCBI BLAST server (<http://www.ncbi.nlm.nih.gov/BLAST>). The search was performed in May 2011. The amino acid sequence of PaaA from *E. coli* K12 was used as a query for a BLASTP search against assembled archaeal genomes (http://www.ncbi.nlm.nih.gov/utils/genom_table.cgi). Unfinished archaeal genome projects were excluded from the BLASTP search. Occurrence and similarity of PaaA was analysed in a total of 87 completely sequenced archaeal genomes. The percentage of archaeal species, which contained homologues, was 10% (9 out of 87 archaeal genomes). Homologues taken into account had an amino acid identity of ≥26% (coverage ≥71%). For construction of a phylogenetic tree, the amino acid sequences of a selection of the bacterial and archaeal PaaA orthologues as well as BoxB¹⁵ orthologues were aligned using CLUSTAL W implemented within MEGA4⁴⁰. The evolutionary history was inferred using the Neighbour-Joining method⁴¹. The phylogenetic tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. Phylogenetic analyses were conducted in MEGA4.

31. Erb, T. J. *et al.* Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. USA* **104**, 10631–10636 (2007).
32. Erb, T. J., Ismail, W. & Fuchs, G. Phenylacetate metabolism in thermophiles: characterization of phenylacetate-CoA ligase, the initial enzyme of the hybrid pathway in *Thermus thermophilus*. *Curr. Microbiol.* **57**, 27–32 (2008).
33. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
34. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
35. Ramos-Vera, W. H., Weiss, M., Strittmatter, E., Kockelkorn, D. & Fuchs, G. Identification of missing genes and enzymes for autotrophic carbon fixation in crenarchaeota. *J. Bacteriol.* **193**, 1201–1211 (2011).
36. Textor, S. *et al.* Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. *Arch. Microbiol.* **168**, 428–436 (1997).

37. Beinert, H. & Thomson, A. J. Three-iron clusters in iron-sulfur proteins. *Arch. Biochem. Biophys.* **222**, 333–361 (1983).
38. Teufel, R., Kung, J. W., Kockelkorn, D., Alber, B. E. & Fuchs, G. 3-hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the Sulfobolales. *J. Bacteriol.* **191**, 4572–4581 (2009).
39. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
40. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
41. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

CAREERS

TURNING POINT Stem-cell biologist carves his own path **p.365**

CAREERS BLOG The latest job discussions and news go.nature.com/ielkfk

 **NATUREJOBS** For the latest career listings and advice www.naturejobs.com

ANDREW HOLBROOK/CORBIS



ENVIRONMENT

Toxic effects

Environmental concerns and more stringent laws are providing opportunities for environmental toxicologists.

BY AMANDA MASCARELLI

When oil began gushing into the Gulf of Mexico after BP's Deepwater Horizon drilling rig exploded in April 2010, politicians and the public were braced for the worst. Oil lapped up on hundreds of kilometres of beach and coastal sands stretching from Florida to Texas, and threatened Louisiana's ecologically sensitive wetlands. Although the wetlands and shorelines faced substantial damage, these regions were spared the brunt of the disaster, the largest marine spill in US history. Instead, much of the oil and gas — and between 2 and 3 million litres of dispersant that BP used to help break up the oil — drifted through the sea in deep underwater plumes, exposing marine organisms, many of which were in the vulnerable reproductive stages of their life cycles, to

pollutants for several months.

Many scientists were alarmed by the unprecedented use of dispersants in the deep ocean. Little was known about how the dispersants and oil would interact, how toxic they were to many marine organisms, how long they would persist in the deep-water environment at potentially damaging concentrations and what effect this would have on the growth and reproduction of the Gulf of Mexico's myriad organisms.

Toxicologists working for government agencies, local universities and private consultants — along with fisheries biologists, oceanographers and ecologists — set to work to find the answer. "It's like CSI meets [Jacques] Cousteau," says Susan Shaw, a marine toxicologist and director of the Marine Environmental Research Institute, a non-profit institution based in Blue Hill, Maine.

During the gulf spill, which took nearly three months to contain, Shaw helped to analyse the effects of the oil and dispersants on marine animals. Later she took part in a scientific working group, assembled by the US Department of the Interior, that was tasked with assessing the effects of the spill on fisheries, wildlife and human health, and with making policy recommendations. This real-world application of her work is what drew Shaw to environmental toxicology 30 years ago. She still finds it deeply gratifying.

Environmental toxicologists aim to understand the effects of chemicals and physical agents — such as dust, mould, cigarette smoke, vehicle exhaust, pesticides, radiation and heavy metals — on both human health and ecosystems. Some work in risk assessment, relying heavily on statistical computations and mathematical modelling to determine whether an agent is likely to be hazardous at a certain level of exposure. Others are mainly chemists or biologists who tackle issues such as how pollutants travel and disperse in air and water, and how these agents affect humans, animals and ecosystems. "It takes the team of those people to put the whole picture together," says Jennifer Field, an environmental chemist at Oregon State University in Corvallis, who traces pharmaceutical products and illicit drugs in wastewater systems as an alternative indicator of drug-use trends.

In the 1970s and 1980s, the field of environmental toxicology expanded rapidly, especially in the United States, triggered mainly by the 1962 publication of Rachel Carson's *Silent Spring* — an exposé of the path that pollutants, particularly pesticides, take through the environment — and by the establishment of the US Clean Water Act and an array of air-quality laws. Although the field has not been immune to the economic depression of the past few years, toxicologists are in demand in government, academia, private industry and consultancies across the United States and internationally. In the United States, toxicologists held around 117,000 jobs in 2010, up 3.5% from 2008, according to an analysis of US Bureau of Labor Statistics by members of the Society of Toxicology in Reston, Virginia, an organization representing scientists from academia, business and government, and presented as a poster at the their 2011 Annual Meeting. The analysis predicts that the number of US toxicologists will grow to some 153,000 by 2018.

New legislation could fuel a greater ▶



Beijing's Forbidden City surrounded by smog at midday.

► demand. In the European Union, for example, the Registration, Evaluation, Authorisation and Restriction of Chemical Substances (REACH) law took effect in June 2007. Phased in over 11 years, REACH aims to create a standardized, rigorous approach to testing and evaluating chemicals before they reach the market. The law will create jobs for toxicologists across all sectors, including academia, government, business and consulting firms that will require

a blend of technical and managerial skills, says Mark Crane, an environmental toxicologist and founder of WCA Environment, an environmental consultancy based in Faringdon, UK. They will be needed to help to find efficient and effective ways to meet the often complex requirements of REACH, says Crane. Knowledge of how the chemical industry and its supply chains work will be a big advantage. "You need to be able to speak knowledgeably to a client, to give them strategic as well as tactical advice," he says. For instance, to help a client meet REACH mandates, it is important to be able to understand the company's specific needs, such as whether the regulations require animal testing — which can be a lengthy and costly process — or if there is a more efficient method to satisfy the regulations.

In fact, there may not be enough well-trained toxicologists to fulfil the demands of REACH in the public and private sector in Europe in the coming years, says Alan Boobis, a toxicologist at Imperial College London. Others have suggested the same, including in a 2007 European Commission report (see go.nature.com/zbtkdi), which noted the lack of toxicology training programmes. The problem, says Boobis, is that toxicologists with the integrated training required for

risk assessment are close to retirement. Furthermore, many universities in Europe offer programmes that do not provide the broader scope of training needed for the applied science of toxicology. "With greater specialization, we are losing this expertise in Europe, and without an adequate career track there is little incentive to pursue such a career."

There are, however, some notable exceptions, including Britain's Cranfield University, which provides tailored postgraduate programmes that allow students to combine policy and applied work, says Crane. He also points to the University of Insubria in Varese, Italy, which offers a doctoral programme in environmental chemistry and ecotoxicology. The Netherlands has a handful of universities that offer good toxicology programmes, including Wageningen University and Utrecht University.

Toxicologists generally acquire expertise in one or two areas, focusing on, for instance, mathematical modelling and biostatistics, epidemiology, molecular biology, environmental chemistry or pharmacology. Most toxicology jobs require PhD training, although there are opportunities for master's students in government and elsewhere. Some of the most important skills that graduates entering the field can bring are social: understanding their clients' needs and collaborating with other experts. "The people who are going to be successful are people who have that deep expertise in their own field yet have the professional capabilities to collaborate, and to open their mind and to really listen," says Dave Stone, director of the US National Pesticide Information Center, a cooperative between the Oregon State University and the US Environmental Protection Agency in Washington DC.

The United States has at least 125 toxicology programmes, some of which concentrate on a specific subfield such as ecotoxicology, inhalation toxicology, metals toxicology or policy. Indiana University in Indianapolis,

for example, is a centre of ecotoxicology and chemical carcinogenesis; Michigan State University in East Lansing focuses on food toxicology; and The University of Rochester in New York specializes in the effects of toxic chemicals on bones. George Gray, a risk analyst at George Washington University in Washington DC, says that the university's programme emphasizes policy, producing competitive graduates. "It makes them attractive to organizations when they understand not just the science but how it ends up being used," says Gray. A great toxicology programme will "allow students to work with professors in a laboratory setting, have the opportunity to contribute to publications as co-authors, and take a wide range of courses in toxicology, pharmacology, chemistry and ecotoxicology," says Margaret Whittaker, managing director and chief toxicologist at ToxServices, a consulting firm based in Washington DC that assesses health risks from chemicals.

REAL-WORLD PROBLEMS

In Europe, REACH puts the burden of proof-of-safety for chemicals on the chemical industry itself; this is not the case in the United States, where advocates have been fighting for an overhaul of existing US legislation for years. The US Toxic Substances Control Act of 1976 is widely viewed as outdated and fractured, even by members of the chemical industry.

"The US law currently takes the view that chemicals are innocent until proven guilty and REACH takes the opposite view — that chemicals are suspect unless the companies making them and using them can prove that they're safe," says Monica Becker, a sustainability consultant based in Rochester, New York. If the burden of proof were to shift, says Becker, more chemical testing and expert interpretation of toxicity data would be required, which would create a demand for toxicologists. But California is already preparing to implement regulations that will require consumer product manufacturers to investigate whether their products contain 'chemicals of concern', and suggest substitutes. Whittaker expects these regulations to create many private-sector jobs for toxicologists.

Legislation adopted in China in 2010, known as China REACH, is similar to that in Europe



Erica Holloman is hoping to make a difference.

and will drive the country's demand for toxicologists, says Kenneth Leung, an aquatic toxicologist at the University of Hong Kong. He says that the number of industry and academic postgraduate and postdoctoral jobs in the field is growing and this is likely to continue as China attempts to meet demands for efficiency, cleaner air and water, and action on addressing climate change.

"North America confronted decades ago the problems now emerging in Asia, so individuals trained in environmental toxicology can make a substantial contribution in Asia," says Michael Newman, an ecotoxicologist at the Virginia Institute of Marine Science in Gloucester Point. China has a number of well-known training programmes, including a programme at the Guangzhou Institute of Geochemistry of the Chinese Academy of Sciences, known for its monitoring of environmental pollution, and the Research Center for Eco-Environmental Sciences, also part of the Chinese Academy of Sciences in Beijing, which is working to develop an early-warning system for environmental contamination.

Globally, the field is moving away from a focus on individual chemicals and towards understanding the cocktail of chemicals that pervade the environment, says Linda Birnbaum, head of both the National Institute of Environmental Health Sciences in Research Triangle Park, North Carolina and the US National Toxicology Program, a federal interagency programme. "People are finally beginning to realize that nobody is exposed to one chemical at a time," says Birnbaum. "We live in a soup. We've got to go beyond looking at one chemical or one exposure at a time and start looking in a more integrated fashion." In the future, toxicologists will need to consider cumulative effects of many exposures and work as part of multidisciplinary research teams.

Most environmental toxicologists say they are highly satisfied with their work, partly because they can apply their expertise to pressing, real-world problems. Erica Holloman, who recently completed her PhD in toxicology at the Virginia Institute of Marine Science, is building on her doctoral work, assessing the effects of mercury in seafood and other pollutants in a low-income community in Newport, Virginia, through a community grant from the US Environmental Protection Agency. "I wanted to see my work make a difference in everyday life," says Holloman. "I found this was a way that my passion for science and the research could lead to something that was dear to my heart, and that I could really have an impact in." ■

Amanda Mascarelli is a freelance writer based in Denver, Colorado.

TURNING POINT

Sean Bendall

Sean Bendall, a postdoctoral fellow in stem-cell and cancer biology at Stanford University in Palo Alto, California, won the Dale F. Frey Award for Breakthrough Scientists from the Damon Runyon Cancer Research Foundation in New York on 9 January 2012.

How did you decide where to do your PhD?

After earning my bachelor's degree in biochemistry and microbiology at the University of Victoria in Canada, I decided to work for a year at a proteomics facility while I applied to graduate programmes. I was accepted to some very prestigious places, including the University of Oxford, UK and the Institute for Systems Biology in Seattle, Washington. The decision was tough because the research project that excited me the most was not happening at Oxford or in Seattle. I wanted to work on the mechanism that makes an embryonic stem cell continue to be a stem cell and stops it from differentiating into another type of cell, and I was able to do this at the University of Western Ontario in London, Canada.

Was it a good decision?

I definitely do not regret deciding to follow my passion. My stem-cell work ended up making a big splash in *Nature* (S.C. Bendall *et al.* *Nature* **448**, 1015–1021; 2007), even though it took a while to get the paper out. I think if you could tell any PhD student that their project would culminate in a high-profile paper in a top journal, they would consider that the best-case scenario. As it was, I left little bits of my soul behind on a long journey to get to that end point.

What did you learn while writing that paper?

How to tell the story better. Like many academics, I was blind when it came to my own research. I had to find a way to write that laid out how my observations could help to guide science in a new direction, rather than simply saying the research so far had been wrong.

How many iterations did the paper undergo?

Several. The first version was two stories — the method development and the actual stem-cell analyses — that were poorly tied together. We sent it to one journal, not *Nature*, and the reviews were not good. But the reviewers did suggest experiments that could strengthen the paper, which helped us to describe, in a new version, how some of the pathways that we had thought were most important in stem-cell cultures were not even acting on the stem cells, but on the support cells. Once I combined my



work with a colleague's, a bigger story emerged of how growth factors help to control how human embryonic stem cells differentiate into, say, bone marrow or skin.

Are there downsides to training at the cutting edge of science?

There can be. In 2008, next-generation flow cytometers were so new that I was almost a year into my postdoc before we got the instrument in the laboratory. Until then, I had to send my samples to the manufacturer in Toronto. Still, once we got everything running well, we were able to quickly publish a paper that illustrated the technology's capabilities and potential. Every project in our lab is now using this technology. It is an exciting time.

Has your ongoing success delayed your efforts to seek a permanent position?

Yes. I haven't applied for any jobs yet because everything is going so well, and I want to focus on the research. But I realize that I need to put myself on the job market and that now is probably best because my work is going so well. Receiving the US\$100,000 Dale F. Frey award is a nice nest egg for when I get my own lab.

What is your secret for success?

Research topics in the top journals are probably not the best PhD or postdoc topics. They already have enough people working on them. I like finding the empty space in research in which no one else is poking around. And my publication record is good partly because I work in multidisciplinary groups. Some researchers think that they should do everything on their own, but that can take so much longer, and a collaborative project might also result in a better interpretation of the data because the case has to be proved to many disciplines. ■

INTERVIEW BY VIRGINIA GEWIN

and will drive the country's demand for toxicologists, says Kenneth Leung, an aquatic toxicologist at the University of Hong Kong. He says that the number of industry and academic postgraduate and postdoctoral jobs in the field is growing and this is likely to continue as China attempts to meet demands for efficiency, cleaner air and water, and action on addressing climate change.

"North America confronted decades ago the problems now emerging in Asia, so individuals trained in environmental toxicology can make a substantial contribution in Asia," says Michael Newman, an ecotoxicologist at the Virginia Institute of Marine Science in Gloucester Point. China has a number of well-known training programmes, including a programme at the Guangzhou Institute of Geochemistry of the Chinese Academy of Sciences, known for its monitoring of environmental pollution, and the Research Center for Eco-Environmental Sciences, also part of the Chinese Academy of Sciences in Beijing, which is working to develop an early-warning system for environmental contamination.

Globally, the field is moving away from a focus on individual chemicals and towards understanding the cocktail of chemicals that pervade the environment, says Linda Birnbaum, head of both the National Institute of Environmental Health Sciences in Research Triangle Park, North Carolina and the US National Toxicology Program, a federal interagency programme. "People are finally beginning to realize that nobody is exposed to one chemical at a time," says Birnbaum. "We live in a soup. We've got to go beyond looking at one chemical or one exposure at a time and start looking in a more integrated fashion." In the future, toxicologists will need to consider cumulative effects of many exposures and work as part of multidisciplinary research teams.

Most environmental toxicologists say they are highly satisfied with their work, partly because they can apply their expertise to pressing, real-world problems. Erica Holloman, who recently completed her PhD in toxicology at the Virginia Institute of Marine Science, is building on her doctoral work, assessing the effects of mercury in seafood and other pollutants in a low-income community in Newport, Virginia, through a community grant from the US Environmental Protection Agency. "I wanted to see my work make a difference in everyday life," says Holloman. "I found this was a way that my passion for science and the research could lead to something that was dear to my heart, and that I could really have an impact in." ■

Amanda Mascarelli is a freelance writer based in Denver, Colorado.

TURNING POINT

Sean Bendall

Sean Bendall, a postdoctoral fellow in stem-cell and cancer biology at Stanford University in Palo Alto, California, won the Dale F. Frey Award for Breakthrough Scientists from the Damon Runyon Cancer Research Foundation in New York on 9 January 2012.

How did you decide where to do your PhD?

After earning my bachelor's degree in biochemistry and microbiology at the University of Victoria in Canada, I decided to work for a year at a proteomics facility while I applied to graduate programmes. I was accepted to some very prestigious places, including the University of Oxford, UK and the Institute for Systems Biology in Seattle, Washington. The decision was tough because the research project that excited me the most was not happening at Oxford or in Seattle. I wanted to work on the mechanism that makes an embryonic stem cell continue to be a stem cell and stops it from differentiating into another type of cell, and I was able to do this at the University of Western Ontario in London, Canada.

Was it a good decision?

I definitely do not regret deciding to follow my passion. My stem-cell work ended up making a big splash in *Nature* (S.C. Bendall *et al.* *Nature* **448**, 1015–1021; 2007), even though it took a while to get the paper out. I think if you could tell any PhD student that their project would culminate in a high-profile paper in a top journal, they would consider that the best-case scenario. As it was, I left little bits of my soul behind on a long journey to get to that end point.

What did you learn while writing that paper?

How to tell the story better. Like many academics, I was blind when it came to my own research. I had to find a way to write that laid out how my observations could help to guide science in a new direction, rather than simply saying the research so far had been wrong.

How many iterations did the paper undergo?

Several. The first version was two stories — the method development and the actual stem-cell analyses — that were poorly tied together. We sent it to one journal, not *Nature*, and the reviews were not good. But the reviewers did suggest experiments that could strengthen the paper, which helped us to describe, in a new version, how some of the pathways that we had thought were most important in stem-cell cultures were not even acting on the stem cells, but on the support cells. Once I combined my



work with a colleague's, a bigger story emerged of how growth factors help to control how human embryonic stem cells differentiate into, say, bone marrow or skin.

Are there downsides to training at the cutting edge of science?

There can be. In 2008, next-generation flow cytometers were so new that I was almost a year into my postdoc before we got the instrument in the laboratory. Until then, I had to send my samples to the manufacturer in Toronto. Still, once we got everything running well, we were able to quickly publish a paper that illustrated the technology's capabilities and potential. Every project in our lab is now using this technology. It is an exciting time.

Has your ongoing success delayed your efforts to seek a permanent position?

Yes. I haven't applied for any jobs yet because everything is going so well, and I want to focus on the research. But I realize that I need to put myself on the job market and that now is probably best because my work is going so well. Receiving the US\$100,000 Dale F. Frey award is a nice nest egg for when I get my own lab.

What is your secret for success?

Research topics in the top journals are probably not the best PhD or postdoc topics. They already have enough people working on them. I like finding the empty space in research in which no one else is poking around. And my publication record is good partly because I work in multidisciplinary groups. Some researchers think that they should do everything on their own, but that can take so much longer, and a collaborative project might also result in a better interpretation of the data because the case has to be proved to many disciplines. ■

INTERVIEW BY VIRGINIA GEWIN

THE PI THAT WASN'T ROUND

How to square a circle.

BY DAN ERLANSON

“**4**.0!” Amisha shouted. The other students laughed.

James gave them a look to keep them quiet. “Very funny, Amisha. Try again.”

“But I’m serious, Mr Wong — look!”

Amisha wasn’t the brightest kid in fourth grade, but she was hard-working, and she certainly wasn’t the class clown. James walked over to investigate. Along with everyone else, Amisha had used a compass to draw a circle on a piece of paper and then used a string and a ruler to measure its circumference and diameter. His class was celebrating March 14, Pi Day, and the students had been reading off their measurements. A ten-year-old and a piece of string can get you only so many significant figures, but Amisha was way outside the margin of error. James watched as Amisha remeasured her circle: 41.6 centimetres in circumference, 10.4 centimetres in diameter. The ratio expected for a square, not a circle.

“Try drawing another,” James suggested. Amisha took a fresh piece of paper, carefully drew, and then measured her new circle: 23.1 centimetres in circumference, 7.4 centimetres in diameter. “3.1?” Amisha asked, a little uncertain. She picked up the string and turned back to her original drawing.

“Perfect!” James said. “I’ll keep this other one.”

After classes ended, James closely studied Amisha’s drawing: just a plain sheet of paper with a circle filling up about a quarter of the page, slightly off-centre. He carefully measured it: 41.6 centimetres in circumference, 10.4 centimetres in diameter: $C/d = 4.0$. Did it feel heavier than a normal sheet of paper? Was there a certain luminescence to it?

He took it down the hall to the staff room and made a photocopy. It looked identical, but James measured it again: 32.8 centimetres in circumference this time, still 10.4 centimetres in diameter: $C/d = 3.15$. He scanned the drawing, saved the image, printed it and measured again: 10.4 centimetres in diameter, 32.8 centimetres in circumference. Whatever was strange about Amisha’s circle, it couldn’t be copied.

All evening he thought about the non-circular circle. His aborted PhD thesis was on a hypothetical particle that could cause spatial distortions if it interacted with ordinary matter. Could one of these ‘strangelets’

have collided with Amisha’s paper?

James needed someone to help him think this through. And although he dreaded reconnecting, he realized that the best person to turn to was his old adviser, Professor Myers. He waited until after dinner — and a stiff drink — before making the call. Of course Myers was still in his office.

“It’s been a while I know, but I have something to show you. Something I think you’ll find interesting.”

There was a pause, then: “James? Gotten bored with the brats?”

“Can I come over?”

“Still remember how to find the department? Give me a call when you arrive and I’ll buzz you in.”

On his drive to the university James thought about the last conversation he had had with Myers. It was an early spring evening like this one, and James was clearing out his desk in the physics department. He had already told Myers about his decision to leave, and Myers had said little, but didn’t hide his disappointment. As James was taking down the last of his photographs, he felt someone approach from behind.

“So what are you going to do?” Myers asked.

“I don’t know; I was thinking of teaching elementary school.”

“Teach!” Myers yelled. “Kids are even dumber than graduate students!” Then, in a lower voice: “You were the brightest student in your class — why do you want to throw that away?”

Because I was becoming too much like you, a grasping egotist with no interest in anything or anyone except his own reputation, James thought furiously, but he replied simply: “I’ve gotta go.”

James had continued reading about physics in the years since. The work he had been doing remained untested and seemingly untestable, and

Myers sank into deeper obscurity, even among physicists, although he apparently remained as

abrasive as ever. James had not been the first student to leave without graduating, and he certainly wasn’t the last.

The building still had the same faint electrical odour that James remembered. Myers’ office door was open as usual so he could keep track of his group. You quickly learned to arrive early or leave late — preferably both.

“So, what have you got? Need help grading?”

James couldn’t help smiling. “In a sense, yes. You remember the project I was working on? Calculating the cross-section of strangelets —”

“Of course I remember — damn shame you didn’t keep at it.”

“I didn’t come here to argue about the past,” James managed evenly. He showed Myers the drawing and explained his observations, then watched as his former adviser measured it again and again before getting up and shutting the door to his office.

“Do you realize what this may mean? This could be the first experimental evidence for my theories.”

“Our theories,” James reminded him.

“Yes of course, you’ll get credit for finding this. But let me keep it for a while — I want to see if temperature or magnetic fields have an effect on the measurements.”

There was something James didn’t like in Myers’ voice. “That’s OK, I’ll hold onto it for now,” he said, picking up the paper.

“Wait, you don’t realize how important this is!” Myers snatched at the drawing and pulled.

There was an audible rip, and the room filled with a brilliant emerald light. The paper seemed to shrivel in upon itself, becoming smaller and smaller until it disappeared altogether with a soft sizzling sound. The preternatural light vanished, and the two men were left staring at the empty space between them. ■

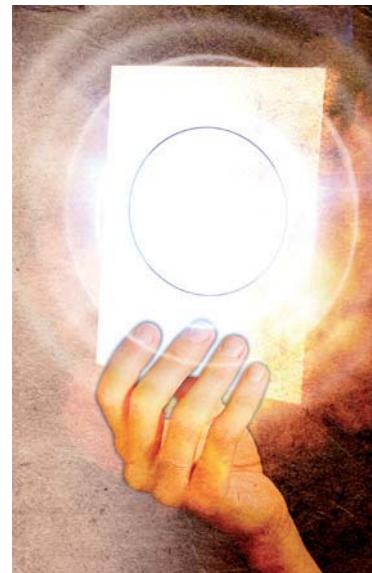
© NATURE.COM

Follow Futures on

Facebook at:

go.nature.com/mtoodm

Dan Erlanson is a chemist who mostly enjoyed graduate school and is still on good terms with his PhD adviser.



MEGF10 and MEGF11 mediate homotypic interactions required for mosaic spacing of retinal neurons

Jeremy N. Kay¹, Monica W. Chu¹ & Joshua R. Sanes¹

In many parts of the nervous system, neuronal somata display orderly spatial arrangements¹. In the retina, neurons of numerous individual subtypes form regular arrays called mosaics: they are less likely to be near neighbours of the same subtype than would occur by chance, resulting in ‘exclusion zones’ that separate them^{1–4}. Mosaic arrangements provide a mechanism to distribute each cell type evenly across the retina, ensuring that all parts of the visual field have access to a full set of processing elements². Remarkably, mosaics are independent of each other: although a neuron of one subtype is unlikely to be adjacent to another of the same subtype, there is no restriction on its spatial relationship to neighbouring neurons of other subtypes⁵. This independence has led to the hypothesis that molecular cues expressed by specific subtypes pattern mosaics by mediating homotypic (within-subtype) short-range repulsive interactions^{1,4–9}. So far, however, no molecules have been identified that show such activity, so this hypothesis remains untested. Here we demonstrate in mouse that two related transmembrane proteins, MEGF10 and MEGF11, have critical roles in the formation of mosaics by two retinal interneuron subtypes, starburst amacrine cells and horizontal cells. MEGF10 and 11 and their invertebrate relatives *Caenorhabditis elegans* CED-1 and *Drosophila* Draper have hitherto been studied primarily as receptors necessary for engulfment of debris following apoptosis or axonal injury^{10–14}. Our results demonstrate that members of this gene family can also serve as subtype-specific ligands that pattern neuronal arrays.

The retina contains over 70 neuronal subtypes, divided into broad categories of photoreceptors, interneurons and retinal ganglion cells¹⁵. To seek molecules involved in cell–cell recognition events during retinal circuit assembly, we purified 13 subtypes of retinal neurons from transgenic mice and used microarrays to inventory the genes they expressed^{16,17}. We collected cells at postnatal day (P)6, a time at which synapse formation and mosaic refinement are underway. From this data set we identified genes selectively expressed by specific subtypes, including starburst amacrine cells (SACs), an interneuronal subtype that has critical roles in motion detection¹⁸. The approximately 100 genes that met our criteria for selective SAC expression included most known SAC markers as well as potential new markers (Fig. 1a and Supplementary Table 1). Thirty-one of the novel genes were tested by *in situ* hybridization or immunohistochemistry in combination with SAC markers such as choline acetyltransferase (ChAT) or calbindin¹⁹. This secondary screen yielded 26 genes (Supplementary Table 1), among which were two homologues, *Megf10* and *Megf11*. These genes encode transmembrane proteins with multiple epidermal growth factor (EGF)-like domains, a single membrane-spanning region, and a cytoplasmic domain with several binding sites for signal transduction components^{10,13} (Supplementary Fig. 1). At P5–6, both genes were strongly expressed in SACs (Fig. 1a–c). Both were also expressed in horizontal cells (HCs), which were not part of the data set.

SACs are present in both the inner nuclear and ganglion cell layers, and form independent mosaics in each⁵. These mosaics develop during late embryonic stages, as newborn SACs migrate from the site of their

birth, the outer neuroblast layer, to their final laminar locations. SACs begin to exhibit mosaic spacing upon arrival at their destinations, presumably due to contacts with their homotypic neighbours^{8,9}. As

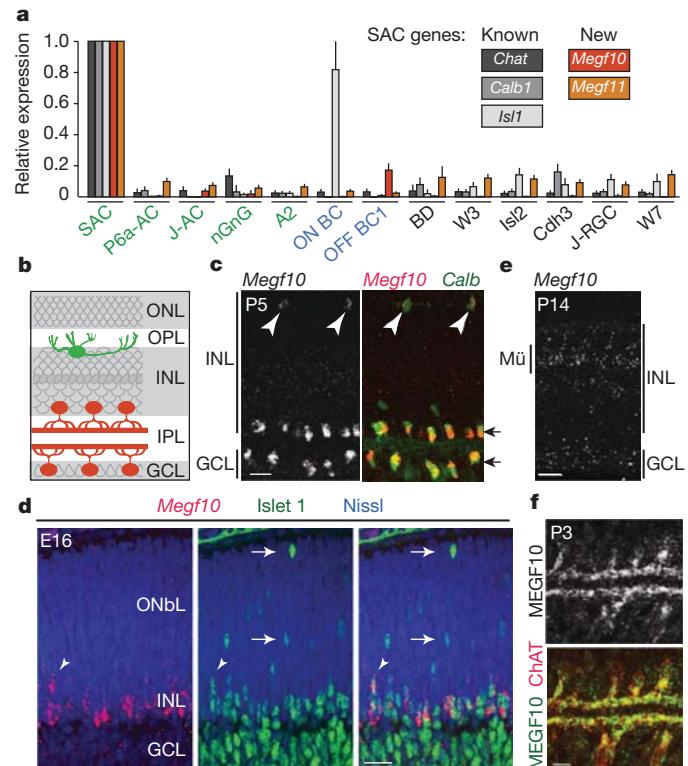


Figure 1 | Expression of *Megf10* and *Megf11* in SACs and HCs. **a**, Relative expression level of *Megf10*, *Megf11* and the known SAC markers *Chat*, *Calb1* (encoding calbindin) and *Isl1* (islet 1) in 13 amacrine (green), bipolar (blue) and retinal ganglion cell (black) subtypes analysed with microarrays. Level in SACs set to 1 for each gene. *Isl1* was also detected in ON bipolar cells as previously reported²¹. Abbreviations for subtypes are defined in Supplementary Table 2. **b**, Schematic of retina. INL, inner nuclear layer with horizontal, bipolar, amacrine, and Müller glial cells; IPL, inner plexiform layer with synapses among bipolar, amacrine and ganglion cells; GCL, ganglion cell layer including displaced amacrine cells; ONL, outer nuclear layer containing photoreceptors; OPL, outer plexiform layer with photoreceptor synapses. SACs (red), HCs (green) and Müller cells (dark grey row) are indicated. **c**, *In situ* hybridization for *Megf10* (left panel; red in right panel) combined with anti-calbindin immunohistochemistry (green in right panel) at P5 shows expression in SACs (black arrows) and HCs (arrowheads). **d**, *Megf10* (red) in E16 retina. Islet 1 (green) marks SACs migrating through the outer neuroblast layer (ONBL) and in the INL⁹. *Megf10* (red) is expressed by migrating SACs as they arrive in the INL (arrowhead), but not at earlier stages of their migration through the ONBL (arrows). **e**, *Megf10* expression appears in Müller glia (Mü) and is lost from SACs and HCs by P14. **f**, Double-label immunostaining for MEGF10 (green) and ChAT (red). MEGF10 protein localizes to the somata and processes of developing SACs (also see Supplementary Fig. 1). Scale bars, 20 μ m (c–e) or 10 μ m (f).

¹Center for Brain Science and Department of Molecular and Cellular Biology, Harvard University, 52 Oxford Street, Cambridge, Massachusetts 02138, USA.

new SACs are added to the array, local cellular rearrangements maintain mosaic spacing^{1,4,9}. Mosaic spacing is maintained even as SAC dendrites grow to overlap with those of their neighbours²⁰; thus, mosaicism is distinct from the phenomenon called tiling, which minimizes dendritic overlap¹⁵. Co-staining at embryonic day (E)16 with the early SAC marker islet 1 (refs 9, 21) demonstrated that newborn SACs activated *Megf10* expression as they finished their migration and became integrated into mosaics (Fig. 1d). *Megf10* expression persisted in SACs from E16 to the first postnatal week, and began in HCs at P0 (Supplementary Fig. 1 and data not shown). In the second postnatal week, *Megf10* was downregulated in neurons but appeared in Müller glia (Fig. 1e), consistent with previous reports that *Megf10* is expressed by brain glia²². To determine the subcellular localization of MEGF10 in neurons, we generated an antibody to recombinant protein (Supplementary Fig. 1). MEGF10 was present both on the somata of SACs and HCs and on their processes (Fig. 1f and Supplementary Fig. 1). As expected, immunoreactivity levels on SACs and HCs were highest during the first postnatal week, and then declined (data not shown). Thus, *Megf10* is expressed by SACs and MEGF10 protein is present on SAC processes during the time that mosaics form. Because *Megf11* expression was not observed in retina until after SAC mosaics had formed (see later), we focused first on *Megf10*.

To investigate whether MEGF10 is required for SAC development, we generated mutant mice (Supplementary Fig. 2). *Megf10* mutants were viable and fertile and their retinas exhibited no gross abnormalities.

Mutant SACs migrated to the inner nuclear and ganglion cell layers as in controls, they were present in normal numbers, and their dendrites projected to appropriate sublaminae of the inner plexiform layer (Fig. 2a–c). Examination of whole mounts showed, however, a marked loss of regular spacing among SAC somata, suggesting that their mosaic arrangement had been disrupted (Fig. 2a and Supplementary Fig. 3).

To assess the degree to which SAC mosaics were disrupted, we measured the exclusion zone—the region surrounding each SAC in which other SACs are rarely found. This parameter is calculated from the density recovery profile, a plot of cell density as a function of distance from each SAC in the array²³ (Fig. 2d). The SAC exclusion zone in mutants was smaller than in wild-type littermates, and was approximately equal to the diameter of a SAC soma (Fig. 2e). As the only limitation on proximity was soma size, SACs seem to be positioned randomly in *Megf10* mutants. This conclusion was supported by two additional measurements of spatial order: the packing factor²³, another index of regularity calculated from the density recovery profile; and an independent measure, the Voronoi domain regularity index, which quantifies variations in the area of the territories nearest to each cell in an array^{20,24}. In each case the index calculated for SAC arrays in *Megf10* mutants was similar to that measured for computer-generated random arrays, whereas SAC arrays from wild-type littermates were highly ordered (Fig. 2f, g); heterozygotes showed a mild disorganization (Supplementary Fig. 4). Together these results suggest that MEGF10 acts in SACs to impose a minimal intercellular spacing; in

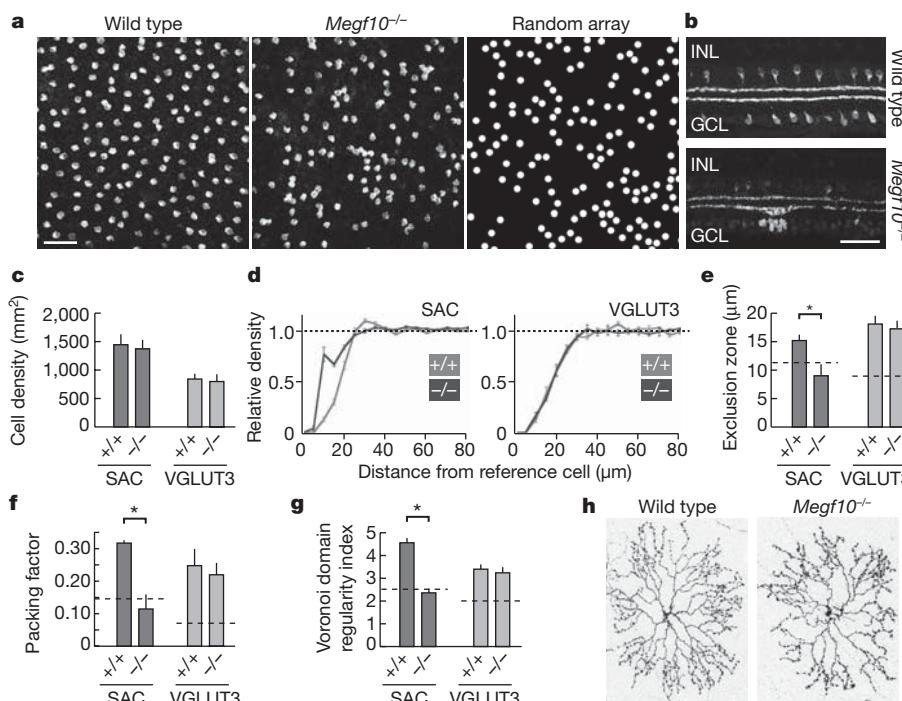


Figure 2 | Loss of SAC mosaic spacing in *Megf10* mutant mice. **a**, SAC mosaic in inner nuclear layer (INL) of wild-type (left) and *Megf10*^{-/-} (centre) retina, revealed by whole-mount staining with anti-ChAT. Wild-type mice have evenly spaced SAC somata, whereas mutants exhibit clumps and gaps similar to those seen in a simulation of a random cellular array (right). See Supplementary Fig. 3 for similar results in ganglion cell layer (GCL) SACs. **b**, ChAT-stained retinal sections from wild-type and *Megf10* mutant animals. Laminar positions of SAC somata and processes are normal in mutants, even in regions where somata are clumped. **c**, Density of SACs and VGLUT3⁺ amacrine cells is similar in wild-type (+/+) and *Megf10* mutant (−/−) retina. **d**, Density recovery profiles (DRPs) for the SAC (INL) and VGLUT3⁺ amacrine arrays. Graphs show the density of cells in a ring of radius x , relative to the density of cells in the image as a whole. Dashed line, DRP of random point array. The exclusion zone characteristic of mosaic spacing is measured as a dip below this line. **e**, Exclusion zone radius measured from **d**. Dashed line, expected result for an array of cells distributed randomly; that is, the diameter of a single cell. Increases above this minimum indicate spatial order. The mutant SAC exclusion zone radius was similar in size to a SAC cell diameter, and was significantly smaller than wild type (* $P < 0.0001$). VGLUT3⁺ amacrine exclusion zones were unaffected. **f**, **g**, SAC packing factor (**f**) and Voronoi domain regularity index (**g**) were significantly lower in *Megf10*^{-/-} mice than in wild-type littermates (* $P < 0.0001$). Dashed line, mean for arrays of cells distributed randomly. Wild-type SACs and VGLUT3⁺ cells of both genotypes were non-randomly arrayed. The SAC array in mutants was not significantly different from random arrays (**f**, $P = 0.16$; **g**, $P = 0.48$). **h**, Morphology of single GCL SACs, labelled with adeno-associated virus driving membrane-targeted Cherry fluorescent protein, showed no gross abnormalities in *Megf10* mutants ($n \geq 8$ cells each genotype). Data from P15 (**a**–**g**) or P80 (**h**) mice. Scale bars, 50 μ m. Error bars, s.e.m.

expected result for an array of cells distributed randomly; that is, the diameter of a single cell. Increases above this minimum indicate spatial order. The mutant SAC exclusion zone radius was similar in size to a SAC cell diameter, and was significantly smaller than wild type (* $P < 0.0001$). VGLUT3⁺ amacrine exclusion zones were unaffected. **f**, **g**, SAC packing factor (**f**) and Voronoi domain regularity index (**g**) were significantly lower in *Megf10*^{-/-} mice than in wild-type littermates (* $P < 0.0001$). Dashed line, mean for arrays of cells distributed randomly. Wild-type SACs and VGLUT3⁺ cells of both genotypes were non-randomly arrayed. The SAC array in mutants was not significantly different from random arrays (**f**, $P = 0.16$; **g**, $P = 0.48$). **h**, Morphology of single GCL SACs, labelled with adeno-associated virus driving membrane-targeted Cherry fluorescent protein, showed no gross abnormalities in *Megf10* mutants ($n \geq 8$ cells each genotype). Data from P15 (**a**–**g**) or P80 (**h**) mice. Scale bars, 50 μ m. Error bars, s.e.m.

its absence SACs assume random positions relative to each other. By contrast, other amacrine cell types examined (VGLUT3- (also known as SLC17A8) and tyrosine hydroxylase-positive) as well as bistratified direction-selective retinal ganglion cells, which are prominent synaptic targets of SACs¹⁸, showed normal mosaic spacing (Fig. 2d–g and Supplementary Fig. 3).

In mice lacking the adhesion molecules DSCAM or DSCAML1, certain retinal cell mosaics form normally, then degrade secondary to hyperfasciculation of their neurites^{25,26}. To investigate whether MEGF10 acts in a similar indirect fashion, we labelled individual SACs in mutants. Lack of MEGF10 had no obvious effect on SAC dendritic morphology (Fig. 2h). Moreover, soma disorganization was already evident by P0 (Supplementary Fig. 5), shortly after mosaics form, indicating that MEGF10 affected formation of SAC mosaics.

Next we asked whether MEGF10 is also essential for formation of the HC mosaic. Loss of MEGF10 had only a modest effect on HC regularity (Fig. 3). We therefore considered that MEGF11 might have a redundant role. *Megf11* was not expressed in embryonic retina, but it appeared in HCs and SACs during the first postnatal week and persisted into adulthood (Fig. 3a–d). Importantly, SAC mosaics begin to form before P0, but HC mosaics are established postnatally^{4,6,9}. We therefore generated *Megf11* mutant mice (Supplementary Fig. 2), which, like *Megf10* mutants, showed no gross retinal abnormalities (Supplementary Fig. 6). Examination of HC arrays in *Megf11* mutants revealed a modest decrease in regularity similar to that in *Megf10* mutants, whereas in *Megf10*^{−/−}; *Megf11*^{−/−} double-mutant animals, the HC mosaic was severely disrupted (Fig. 3e–m and Supplementary Fig. 7). By contrast, SACs were unaffected by loss of *Megf11* function, and were no more affected in double mutants than in *Megf10* single mutants (Supplementary Figs 6 and 8). Thus, *Megf11*

is dispensable for SAC arrangement, but acts together with *Megf10* to shape the HC mosaic.

To elucidate the cellular mechanisms by which *Megf10* and *Megf11* act, we used a gain-of-function approach in which we introduced MEGF10 into the retina by electroporation of plasmid DNA at P0. We first tested the hypothesis that MEGF10 can act as a signal that repels SACs, creating the exclusion zone that defines mosaic spacing. The electroporation method predominantly transfects dividing cells, leading to expression in neurons that exit the cell cycle postnatally, such as bipolar cells, Müller glia, photoreceptors and late-born amacrine cells²⁷. Because SACs and HCs are born embryonically, they are rarely transduced²⁷, allowing us to surround wild-type SACs and HCs with cells ectopically expressing excess MEGF10. Indeed, electroporation of plasmid encoding a fluorescent protein (FP) or a MEGF10–FP fusion produced retinal patches in which a large fraction of neurons, but no SACs or HCs, were FP⁺ (Fig. 4a and Supplementary Fig. 9). Expression of MEGF10–FP (but not FP alone) resulted in exclusion of SACs and HCs from a swath at the edge of the electroporated patch, whereas spacing of these cells was essentially normal in patch centres (Fig. 4a, b and Supplementary Fig. 10). A truncated MEGF10 lacking the cytoplasmic domain produced an identical phenotype (Supplementary Fig. 10), ruling out the possibility that MEGF10 acts indirectly by triggering production of a repellent factor by the transfected cells. Thus, MEGF10 can act as a ligand that signals to SACs and HCs.

Because MEGF10 relatives have been implicated in cell engulfment^{10–13}, we asked whether its overexpression formed an exclusion zone by eliminating cells. We found no evidence for cell death at patch edges (no pyknotic nuclei or activated caspase-3 immunoreactivity; data not shown). Instead, cell density was increased at the outer edge of the cell-free swath, indicating that SACs had exited the patch to create the

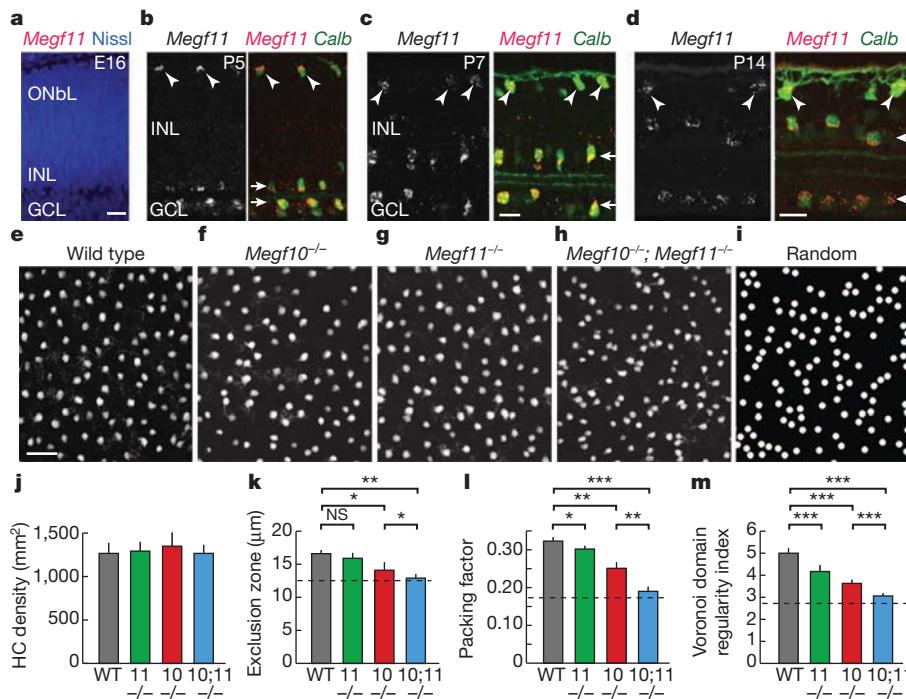


Figure 3 | HC mosaic spacing requires *Megf10* and *Megf11*. **a–d**, *In situ* hybridization for *Megf11* at ages indicated. *Megf11* (red) was not expressed at E16 (a). Calbindin immunostaining (green) labels SACs and HCs, which co-express *Megf11* at P5 (b), P7 (c) and P14 (d). See Fig. 1 for abbreviations. **e–i**, Retinal whole mounts stained for calbindin to reveal the HC array. In wild-type mice (e), HCs are distributed evenly. *Megf10*^{−/−} mutants (f) and *Megf11*^{−/−} mutants (g) show subtle changes in the regular spacing of HCs, whereas double *Megf10*^{−/−}; *Megf11*^{−/−} mutants (h) show marked HC disorganization similar to a simulation of a random HC array (i). **j–m**, Quantification of HC spacing regularity in *Megf10*^{−/−} (red) or *Megf11*^{−/−} (green) single mutants; *Megf10*^{−/−}; *Megf11*^{−/−} double mutants (blue); and wild-type (WT) siblings (grey). In all genotypes, HCs were present at normal density (j) but were less regularly spaced relative to wild type based on exclusion zone radius (k), packing factor (l) and Voronoi domain regularity (m) measurements as in Fig. 2. Double mutants showed significantly less order than single *Megf10*^{−/−} or *Megf11*^{−/−} mutants and approach random arrangement, indicated by dashed lines (k, mean HC soma diameter; l, m, computed values for random arrays). *P* values: **P* < 0.01, ***P* < 0.001, ****P* < 0.0001. NS, not significant. Error bars give s.e.m. Data in e–m from P15 animals. Scale bars, 20 μm (a–d, b, c share scale) or 50 μm (e–i).

Megf11^{−/−} double mutants (blue); and wild-type (WT) siblings (grey). In all genotypes, HCs were present at normal density (j) but were less regularly spaced relative to wild type based on exclusion zone radius (k), packing factor (l) and Voronoi domain regularity (m) measurements as in Fig. 2. Double mutants showed significantly less order than single *Megf10*^{−/−} or *Megf11*^{−/−} mutants and approach random arrangement, indicated by dashed lines (k, mean HC soma diameter; l, m, computed values for random arrays). *P* values: **P* < 0.01, ***P* < 0.001, ****P* < 0.0001. NS, not significant. Error bars give s.e.m. Data in e–m from P15 animals. Scale bars, 20 μm (a–d, b, c share scale) or 50 μm (e–i).

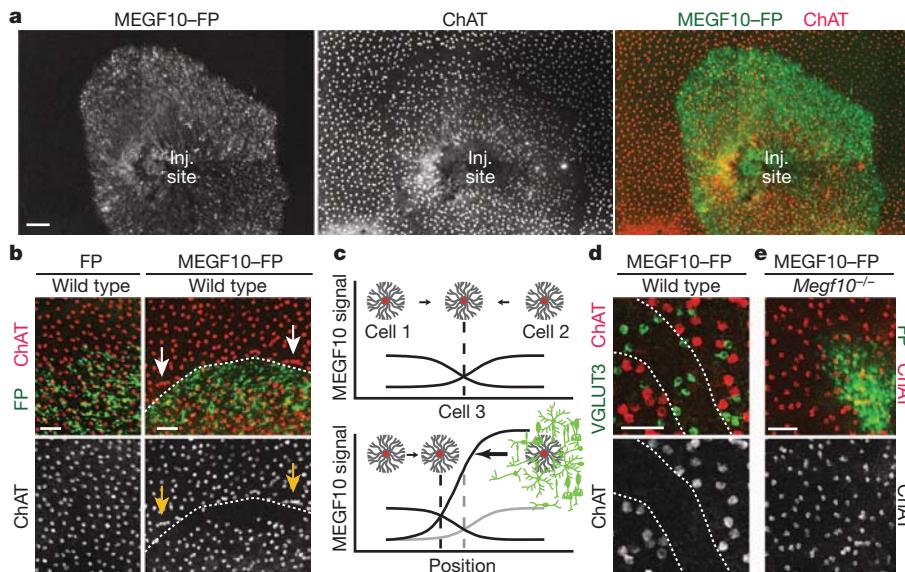


Figure 4 | MEGF10 acts as both ligand and receptor to trigger SAC repulsion. **a**, A retinal patch transfected by electroporation with plasmid encoding MEGF10–FP fusion protein, viewed in flat mount. SAC somata in INL, stained with anti-ChAT, are excluded from a swath at the patch edge. SACs are evenly spaced elsewhere, except where the retina was pierced to inject DNA (Inj. site). **b**, Higher-magnification views of patch edges. FP misexpression (left) did not affect SAC spacing, but MEGF10–FP (right) produced a SAC-free zone just inside the transfected region and induced apparent crowding of SACs immediately outside it (arrows). Dashed line, patch edge. See Supplementary Fig. 11 for quantification of cell distribution at patch edges. **c**, Hypothesis for MEGF10 function based on **a**, **b**. In wild-type retina

SAC-free zone (Fig. 4b and Supplementary Fig. 11). Our interpretation is that SACs and HCs are repelled by MEGF10, but that within MEGF10–FP patches the concentration of MEGF10 is similar in all directions, resulting in no net movement (Fig. 4c). Consistent with this view, SACs and HCs were entirely excluded from small patches (≤ 5 –6 cell diameters; Supplementary Fig. 10). Together these results suggest that MEGF10 can act as a repellent ligand.

We next asked which cells are sensitive to MEGF10 ligand. In *Megf10* mutants, SACs and HCs are selectively affected; this might be because they are uniquely sensitive to MEGF10 or because only these cells encounter endogenous MEGF10 at high concentration. Ectopic expression allowed us to distinguish between these possibilities. None of 13 other amacrine, bipolar or retinal ganglion cell subtypes assayed with cell-type-specific markers was detectably affected by MEGF10–FP (Fig. 4d and Supplementary Table 3). Thus, MEGF10 appears to act as a cell-surface ligand for a receptor specifically expressed by SACs and HCs.

This led us to ask whether MEGF10 might be the MEGF10 receptor. We used cultured epithelial cells (HEK293) to seek evidence for a MEGF10-dependent homotypic interaction. Whereas FP-expressing cells overlapped, MEGF10–FP-expressing cells formed sharp borders with narrow gaps (Supplementary Fig. 12a), consistent with previous results¹⁴. This ‘jigsaw’ pattern reflected suppression or elimination of filopodia at points of intercellular apposition (Supplementary Fig. 12b). Jigsaw formation and loss of filopodia required the MEGF10 cytoplasmic domain and did not occur when a MEGF10–FP-expressing cell contacted an untransfected or FP-transfected cell (Supplementary Fig. 12b, c). Although we have been unable to demonstrate MEGF10 homophilic binding using biochemical methods (data not shown), these results suggest that a MEGF10-containing signalling complex mediates a homotypic interaction resulting in intercellular repulsion.

Finally, we tested the idea that MEGF10 serves as both ligand and receptor *in vivo*. To this end, we electroporated MEGF10 into *Megf10* mutant retina. Mutant SACs did not exit MEGF10–FP patches (Fig. 4e

(top), SACs use MEGF10 as a ligand to signal their location to neighbouring SACs. Cell 3 positions itself at the point where repulsive signals on either side (from Cell 1 and Cell 2) are equal. In MEGF10–FP overexpression (bottom), that location is now outside the patch. **d**, MEGF10–FP transfected whole-mount immunostained for ChAT and VGLUT3. Dashed lines mark SAC-free zone (see Supplementary Fig. 10). VGLUT3⁺ amacrine cells are present in this zone. **e**, MEGF10–FP is incapable of generating a SAC-free zone when misexpressed in *Megf10* mutant retina, indicating that MEGF10 is needed for SAC responses to MEGF10. See Supplementary Fig. 11 for quantification. Scale bars, 100 μ m (**a**); 50 μ m (**b**–**e**). All retinas are from P10–P15 animals.

and Supplementary Fig. 11b), indicating that *Megf10* gene function is required for SACs to respond to MEGF10 repulsive signals. Together these results suggest a model in which SACs and HCs use MEGF10 as part of a receptor complex that detects MEGF10 on their homotypic neighbours. This repulsive signal positions their somata so as to equalize MEGF10 signals on all sides, thereby creating exclusion zones (Fig. 4c).

The phenomenon of retinal mosaicism implies a molecular system for cell-type-specific recognition. Several potential mechanisms have been proposed, based on imaging and computational studies, one being that short-range repulsive signals regulate tangential movements to establish each cell’s territory^{3,4,6–9}. So far, however, direct mediators of this phenomenon have not been described. Here we provide evidence that, for SACs and HCs, mosaic spacing requires repellent homotypic interactions mediated by MEGF10 and 11. Signals initiated by MEGF10/11 in growing neurites could lead to repositioning of the soma within the cytoplasm, perhaps by mechanisms resembling those that direct soma translocation in response to signals at the leading process of migrating neurons²⁸. The finding that MEGF10 and 11 facilitate formation of three independent mosaics (two for SACs and one for HCs) appears at first to contradict the idea that mosaics are independent of each other and must therefore be regulated by distinct molecules^{1,3–5}. However, because the three mosaics occupy distinct cellular planes, they may be exempt from the requirement for molecular individuation. Different molecules are likely to mediate homotypic interactions in other retinal subtypes, some of which may lead to soma translocation and others to the death of cells that violate minimal spacing²⁴.

We note two broader implications of our results. First, Draper/CED-1/MEGF10 homologues have until now been studied predominantly as receptors for cell engulfment^{10–14}. Here we show that they also mediate cell–cell repulsion and can act as ligands as well as receptors, thereby expanding the roles for this gene family. Second, although mosaic arrangements have so far been studied formally only in retina, regularly arranged neuronal arrays are common features of central neural

organization^{1,29}. Mechanisms similar to those described here could be involved in promoting this regularity.

METHODS SUMMARY

Retinal neurons expressing fluorescent proteins were purified and used to generate amplified RNA for hybridizing Affymetrix microarrays as described previously^{16,17}. SAC-specific genes were identified using dChip software. *Megf10* and *Megf11* mutant mice were produced from constructs generated by the Knockout Mouse Project and European Conditional Mouse Mutagenesis program (EUCCOMM)³⁰. Histological methods^{16,17} and methods for electroporation of plasmid DNA *in vivo*^{17,27} were described previously. HEK293 cells (ATCC) were cultured and transfected by standard methods.

For analysis of spatial statistics, we sampled a 635.9 μm square at 3–4 locations per retina. X – Y cell coordinates, marked manually, were used to calculate Voronoi domain areas (Fiji) or DRP statistics such as the effective radius (that is, exclusion zone) and packing factor (WinDRP). We generated random arrays matched in density and soma size to real data; these were analysed in parallel with data from mutants. For measurement of SAC crowding in gain-of-function experiments, the Delaunay triangulation⁸ was used to define each cell's nearest neighbours.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 October 2011; accepted 17 January 2012.

Published online 11 March 2012.

1. Cook, J. E. & Chalupa, L. M. Retinal mosaics: new insights into an old concept. *Trends Neurosci.* **23**, 26–34 (2000).
2. Wässle, H. & Riemann, H. J. The mosaic of nerve cells in the mammalian retina. *Proc. R. Soc. Lond. B* **200**, 441–461 (1978).
3. Eglen, S. J. Development of regular cellular spacing in the retina: theoretical models. *Math. Med. Biol.* **23**, 79–99 (2006).
4. Reese, B. E. & Galli-Resta, L. The role of tangential dispersion in retinal mosaic formation. *Prog. Retin. Eye Res.* **21**, 153–168 (2002).
5. Rockhill, R. L., Euler, T. & Masland, R. H. Spatial order within but not between types of retinal neurons. *Proc. Natl Acad. Sci. USA* **97**, 2303–2307 (2000).
6. Huckfeldt, R. M. *et al.* Transient neurites of retinal horizontal cells exhibit columnar tiling via homotypic interactions. *Nature Neurosci.* **12**, 35–43 (2009).
7. Poché, R. A. *et al.* Somal positioning and dendritic growth of horizontal cells are regulated by interactions with homotypic neighbors. *Eur. J. Neurosci.* **27**, 1607–1614 (2008).
8. Galli-Resta, L. Local, possibly contact-mediated signalling restricted to homotypic neurons controls the regular spacing of cells within the cholinergic arrays in the developing rodent retina. *Development* **127**, 1509–1516 (2000).
9. Galli-Resta, L., Resta, G., Tan, S. S. & Reese, B. E. Mosaics of islet-1-expressing amacrine cells assembled by short-range cellular interactions. *J. Neurosci.* **17**, 7831–7838 (1997).
10. Wu, H.-H. *et al.* Glial precursors clear sensory neuron corpses during development via Jedi-1, an engulfment receptor. *Nature Neurosci.* **12**, 1534–1541 (2009).
11. MacDonald, J. M. *et al.* The *Drosophila* cell corpse engulfment receptor Draper mediates glial clearance of severed axons. *Neuron* **50**, 869–881 (2006).
12. Reddien, P. W. & Horwitz, H. R. The engulfment process of programmed cell death in *Caenorhabditis elegans*. *Annu. Rev. Cell Dev. Biol.* **20**, 193–221 (2004).
13. Hamon, Y. *et al.* Cooperation between engulfment receptors: the case of ABCA1 and MEGF10. *PLoS ONE* **1**, e120 (2006).
14. Suzuki, E. & Nakayama, M. MEGF10 is a mammalian ortholog of CED-1 that interacts with clathrin assembly protein complex 2 medium chain and induces large vacuole formation. *Exp. Cell Res.* **313**, 3729–3742 (2007).
15. Sanes, J. R. & Zipursky, S. L. Design principles of insect and vertebrate visual systems. *Neuron* **66**, 15–36 (2010).
16. Kay, J. N. *et al.* Retinal ganglion cells with distinct directional preferences differ in molecular identity, structure, and central projections. *J. Neurosci.* **31**, 7753–7762 (2011).
17. Kay, J. N., Voinescu, P. E., Chu, M. W. & Sanes, J. R. Neurod6 expression defines new retinal amacrine cell subtypes and regulates their fate. *Nature Neurosci.* **14**, 965–972 (2011).
18. Demb, J. B. Cellular mechanisms for direction selectivity in the retina. *Neuron* **55**, 179–186 (2007).
19. Haverkamp, S. & Wässle, H. Immunocytochemical analysis of the mouse retina. *J. Comp. Neurol.* **424**, 1–23 (2000).
20. Keeley, P. W., Whitney, I. E., Raven, M. A. & Reese, B. E. Dendritic spread and functional coverage of starburst amacrine cells. *J. Comp. Neurol.* **505**, 539–546 (2007).
21. Elshatory, Y. *et al.* Islet-1 controls the differentiation of retinal bipolar and cholinergic amacrine cells. *J. Neurosci.* **27**, 12707–12720 (2007).
22. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
23. Rodieck, R. W. The density recovery profile: a method for the analysis of points in the plane applicable to retinal studies. *Vis. Neurosci.* **6**, 95–111 (1991).
24. Raven, M. A., Eglen, S. J., Ohab, J. J. & Reese, B. E. Determinants of the exclusion zone in dopaminergic amacrine cell mosaics. *J. Comp. Neurol.* **461**, 123–136 (2003).
25. Fuerst, P. G., Koizumi, A., Masland, R. H. & Burgess, R. W. Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature* **451**, 470–474 (2008).
26. Fuerst, P. G. *et al.* DSCAM and DSCAM1 function in self-avoidance in multiple cell types in the developing mouse retina. *Neuron* **64**, 484–497 (2009).
27. Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 16–22 (2004).
28. Solecki, D. J., Model, L., Gaetz, J., Kapoor, T. M. & Hatten, M. E. Par6 α signaling controls glial-guided neuronal migration. *Nature Neurosci.* **7**, 1195–1203 (2004).
29. Budry, L. *et al.* Related pituitary cell lineages develop into interdigitated 3D cell networks. *Proc. Natl Acad. Sci. USA* **108**, 12515–12520 (2011).
30. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Tilton, P. Rogers, J. Couget, and the Harvard Genome Modification Facility for technical assistance; S. Sarin and M. Yamagata for critical discussions; the National Institutes of Health (NS029169 and EY022073 to J.R.S.) and Life Sciences Research Foundation (J.N.K.) for funding.

Author Contributions J.N.K. and J.R.S. designed experiments and wrote the paper. J.N.K. and M.W.C. performed experiments. J.N.K. performed data analysis. J.R.S. supervised the project.

Author Information Data have been deposited at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession code GSE35077. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.R.S. (sanesj@mcb.harvard.edu).

METHODS

Animals. CD1 and C57BL/6 mice were obtained from Charles River and Jackson Labs. All experiments were carried out in accordance with protocols approved by the Harvard University Standing Committee on the Use of Animals in Research and Teaching.

Megf10 mutant mice were produced from a construct provided by the Knockout Mouse Project³⁰ (CHORI). Embryonic stem (ES) cell electroporation and chimaera production were performed by the Harvard Genome Modification Facility. ES cell clones were screened for integration by PCR using primers listed in Supplementary Table 4 (also see Supplementary Fig. 2). *Megf11* mutant mice were generated from gene-targeted ES cells provided by the European Conditional Mouse Mutagenesis program (EUCOMM). In both cases germline transmission was obtained from two chimaeric mice generated from independent ES cell clones; each had indistinguishable phenotypes. The primers used for genotyping *Megf10* and *Megf11* mice are listed in Supplementary Table 4. *Drd4-GFP* mice³¹ were obtained from the Mutant Mouse Regional Resource Center.

Cell sorting and expression profiling. Retinal neurons expressing fluorescent proteins were purified as described^{16,17}. Briefly, P6 retinas were dissociated, live-stained with antibodies recognizing cell surface antigens (if required for purification), and then passed through a flow cytometer (Mo Flo; Dako). Positive cells were then either: (1) plated, fixed, stained with cell-type-specific markers and counted to assess purity; or (2) sorted directly into RNA lysis buffer (PicoPure Kit; MDS). Gene expression was profiled using Affymetrix Mouse 430 2.0 arrays, following two rounds of linear amplification (MessageAmp II; Applied Biosystems). Using these methods we generated a gene-expression database for 13 specific retinal neuron subtypes (five amacrine, two bipolar and six retinal ganglion cell types; see Supplementary Table 2). SAC-specific genes (Supplementary Table 1) were identified using dChip software using fold-change over other cell types as filtering criteria.

Histology and staining. Fixation and preparation of retinal cryosections was performed as described¹⁷. For whole-mount retinal stains, the retina was dissected out of the eyecup in a dish of phosphate buffered saline (PBS), transferred to an Eppendorf tube (up to four retinas per tube), and incubated at room temperature (20–25 °C) for 1 h with agitation in blocking solution (PBS + 0.3% Triton-X-100 + 3% donkey serum; Jackson Immunoresearch). Primary antibody, diluted in blocking solution, was then applied to the sample; incubation was for 6 days at 4 °C with agitation. Following a 2–3 h wash in 2–3 changes of PBS, secondary antibodies (Jackson Immunoresearch) were applied overnight at 4 °C with agitation. After at least 4 h of washing in PBS with agitation, retinas were flattened onto nitrocellulose membranes (Millipore) and mounted on slides (Fluoromount G; Southern Biotech). Images were acquired with a FV1000 confocal microscope (Olympus).

The following antibodies were used for immunostaining: rabbit anti-calbindin (Swant); goat anti-ChAT (Millipore); goat anti-VACHT (Promega); guinea pig anti-VGLUT3 (Millipore); mouse anti-islet 1/2 (clone 39.4D5, Developmental Studies Hybridoma Bank (DSHB)); mouse anti-SYT2 (DSHB); chicken anti-green fluorescent protein (GFP; Aves labs); rabbit anti-DsRed (Clontech); rabbit anti-β-galactosidase³²; mouse anti-glutamine synthetase (BD); sheep anti-tyrosine hydroxylase (Millipore); mouse anti-PKC (Abcam); rabbit anti-DAB1 (Millipore); rabbit anti-CART (Phoenix); mouse anti-Kv4.2 (Neuromabs); rabbit anti-SOX9 (Millipore); rabbit anti-cleaved-caspase 3 (Cell Signaling Technology); rabbit anti-bNOS (Sigma); rabbit anti-EBF3 (ref. 17); rabbit anti-GAD65/67 (Millipore); goat anti-GLYT1 (Millipore); mouse anti-BRN3A (Millipore); mouse SMI-32 (Sternberger). MEGF10 antibody was raised in rabbits against a His-tagged peptide corresponding to amino acids 879–1130 of mouse MEGF10 (Millipore).

In situ hybridization was performed according to previously reported methods¹⁷. To make the template for the *Megf10* antisense probe, we obtained an IMAGE clone (BC075647) containing full-length mouse *Megf10* and linearized with *Clai*. *Megf11* template, in pGEM-T Easy (Promega), was generated by PCR from brain cDNA using primers TGTCTTCCTCTGCAAGCTGCTCT and ATTCCACAAAGT GCCTGGTGAGTGT. Antisense riboprobes were transcribed using the DIG RNA labelling kit (Roche). Calbindin immunostaining was used to label SACs and HCs—at P0–P7 it is a selective marker for these two cell types, showing additional expression in only a small subset of amacrine and retinal ganglion cells (Fig. 1 and Supplementary Fig. 9, and data not shown).

Labelling of single SACs and HCs was performed by intraocular injection of low-titre adeno-associated virus, as described³³.

Quantification of mosaic regularity. For analysis of mutant phenotypes, confocal z-stacks through the GCL and INL were acquired from P14–17 retinas stained with antibodies to ChAT, calbindin, tyrosine hydroxylase, and/or VGLUT3. Each retina was sampled in 3–4 locations (sample size was a square, 635.9 μm on each side). Stacks in which mounting artefacts introduced large local z-axis displacement of the SAC or HC array were not used. We sampled from both central and peripheral

retina; despite differences in cell density between these locations the spatial organization of SAC, HC and VGLUT3 arrays did not appear to differ systematically. We therefore pooled data from central and peripheral retina for subsequent analysis. Using Fiji software, the centre of each cell was marked manually to generate X–Y coordinates. The point array was flattened to a single plane without correction for local curvature, but as we only used very flat images for analysis, the spatial distortions introduced by this procedure were minor. Density recovery profiles (DRPs) were computed from these X–Y coordinates using WinDRP software³⁴. Voronoi domain areas were computed in Fiji and Microsoft Excel from the X–Y coordinate data. SAC, HC and VGLUT3⁺ amacrine cell density and cell diameters were determined from these same images using Fiji to perform measurements.

We calculated three measures of regularity for each image. First, from the DRP, we obtained the effective radius. Effective radius gives the size of the exclusion zone—the zone in which another cell is less likely to be found than would be expected for a random array²³. Second (also from the DRP), we computed the packing factor—a regularity index that ranges between 0 (for a random array) and 1 (for a perfect hexagonal array). Third, we calculated the Voronoi domain regularity index (VDRI) by dividing the mean Voronoi domain area for a given cellular array by the standard deviation of those areas^{20,24,35}.

To ascertain how the measured packing factor and VDRI compared to those that would be observed in cell arrays lacking mosaic spacing, we generated random simulations of HC, SAC, and VGLUT3⁺ amacrine cell arrays. These simulations placed cells randomly in a 635.9 μm square following a Poisson point process, until the density of the array equalled the mean density of the cell type in question. The only constraint on cell location in these simulations was soma diameter, which was calculated for each cell type as described earlier. The soma diameters used to constrain the simulations were 9.0 μm (VGLUT3⁺ amacrine cells), 11.0 μm (SACs) and 12.2 μm (HCs). Programita software³⁶ was used to generate the simulations ($n = 10$ for each cell type). Packing factor and VDRI were then calculated as described above, and the means were plotted as dashed lines in Figs 2 and 3. The VDRIs for our random simulations were similar to those calculated previously³⁵.

For analysis of SAC cell position in MEGF10-misexpressing retina, we acquired images at the edge of FP⁺ or MEGF10–FP⁺ misexpressing patches, and used the above method to obtain X–Y coordinates of all SACs in the field of view. SACs inside the patch, outside the patch, and at the edge of the patch were marked separately. ‘Edge’ cells were defined as those outside the FP⁺ region for which the shortest line drawn from that cell to the edge of the FP⁺ region did not pass the soma of another SAC. To ask whether SACs were present at higher density at patch edges, we calculated the distance from each SAC to its nearest neighbours. Neighbours were defined in an unbiased manner by computing the Delaunay triangulation for the X–Y location data set (Fiji), thereby defining line segments from each cell to its nearest neighbours.

Statistical analysis. For analysis of exclusion zone radius, packing factor, VDRI and cell density, the significance of measured differences between genotypes was evaluated by the Mann–Whitney U test. Sample sizes were ≥ 9 images from ≥ 3 retinas per genotype (SAC and HC analysis) or 6 images from 2 retinas per genotype (VGLUT3 analysis). In double-mutant experiments on HCs, the Holm–Bonferroni correction for multiple comparisons (four different genotypes) was applied when determining significance level. For measurement of soma diameter, sample size was ≥ 150 cells for each genotype.

Gain-of-function experiments. A *Megf10* expression construct was generated by PCR amplifying the open reading frame (ORF), with stop codon deleted, from the IMAGE clone described earlier. The open reading frame, which was predicted to encode a protein equivalent to the *Megf10* RefSeq sequence (NP_115822.1), was TA-cloned into the Gateway entry vector pCR8GW-Topo (Invitrogen). Primers used for cloning were CGATTGTTCTCACAGAACATGGCG and TTGAT GTGATTCACTGCTGCT. A cytoplasmic domain-deletion construct was made by amplifying with reverse primer TGATTCCCTCCTCTGCTT to generate a truncated protein carrying only the first nine amino acids of the intracellular domain. For expression, these constructs were transferred to a Gateway destination vector bearing the ubiquitin-C promoter and an in-frame C-terminal GFP or monomeric Cherry tag. Mouse *Megf11* was also cloned, but pilot experiments in HEK cells suggested a lack of surface expression, so we did not attempt *in vivo* experiments with mouse *Megf11*.

In vivo electroporation was performed as described^{17,27}. Briefly, plasmid DNA (at least 1.5 mg ml⁻¹) was injected into the subretinal space of neonatal mice (4–36 h postpartum), and current pulses (80 V) were applied across the head using paddle electrodes. We obtained identical results in MEGF10-GFP and MEGF10-Cherry misexpression experiments. Sample sizes were > 20 animals each for MEGF10 misexpression and FP controls.

HEK 293 cells were cultured in DMEM with 10% fetal calf serum and transfected using TransIT reagent (Mirus). Cells were counterstained using Alexa

- dye-labelled cholera toxin B subunit ($10\text{ }\mu\text{g ml}^{-1}$; Invitrogen), added to media 30 min before fixation (4% paraformaldehyde/1× PBS for 20 min on ice).
31. Huberman, A. D. *et al.* Genetic identification of an On-Off direction-selective retinal ganglion cell subtype reveals a layer-specific subcortical map of posterior motion. *Neuron* **62**, 327–334 (2009).
 32. Gray, G. E. & Sanes, J. R. Lineage of radial glia in the chicken optic tectum. *Development* **114**, 271–283 (1992).
 33. Hong, Y. K., Kim, I.-J. & Sanes, J. R. Stereotyped axonal arbors of retinal ganglion cell subsets in the mouse superior colliculus. *J. Comp. Neurol.* **519**, 1691–1711 (2011).
 34. Euler, T. WinDRP website. <http://www.mpimf-heidelberg.mpg.de/~teuler/WinDRP/ReadMe.htm> (2003).
 35. Whitney, I. E., Keeley, P. W., Raven, M. A. & Reese, B. E. Spatial patterning of cholinergic amacrine cells in the mouse retina. *J. Comp. Neurol.* **508**, 1–12 (2008).
 36. Wiegand, T. & Moloney, K. A. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* **104**, 209–229 (2004).

Adaptive radiation of multituberculate mammals before the extinction of dinosaurs

Gregory P. Wilson¹, Alistair R. Evans², Ian J. Corfe³, Peter D. Smits^{1,2}, Mikael Fortelius^{3,4} & Jukka Jernvall³

The Cretaceous–Paleogene mass extinction approximately 66 million years ago is conventionally thought to have been a turning point in mammalian evolution^{1,2}. Prior to that event and for the first two-thirds of their evolutionary history, mammals were mostly confined to roles as generalized, small-bodied, nocturnal insectivores³, presumably under selection pressures from dinosaurs⁴. Release from these pressures, by extinction of non-avian dinosaurs at the Cretaceous–Paleogene boundary, triggered ecological diversification of mammals^{1,2}. Although recent individual fossil discoveries have shown that some mammalian lineages diversified ecologically during the Mesozoic era⁵, comprehensive ecological analyses of mammalian groups crossing the Cretaceous–Paleogene boundary are lacking. Such analyses are needed because diversification analyses of living taxa^{6,7} allow only indirect inferences of past ecosystems. Here we show that in arguably the most evolutionarily successful clade of Mesozoic mammals, the Multituberculata, an adaptive radiation began at least 20 million years before the extinction of non-avian dinosaurs and continued across the Cretaceous–Paleogene boundary. Disparity in dental complexity, which relates to the range of diets, rose sharply in step with generic richness and disparity in body size. Moreover, maximum dental complexity and body size demonstrate an adaptive shift towards increased herbivory. This dietary expansion tracked

the ecological rise of angiosperms⁸ and suggests that the resources that were available to multituberculates were relatively unaffected by the Cretaceous–Paleogene mass extinction. Taken together, our results indicate that mammals were able to take advantage of new ecological opportunities in the Mesozoic and that at least some of these opportunities persisted through the Cretaceous–Paleogene mass extinction. Similar broad-scale ecomorphological inventories of other radiations may help to constrain the possible causes of mass extinctions^{9,10}.

Multituberculate mammals were a taxonomically rich^{3,11} and numerically abundant¹² clade that had originated by the Middle Jurassic epoch (approximately 165 million years (Myr) ago) and went extinct in the late Eocene (approximately 35 Myr ago)^{3,11}. They were nearly globally distributed¹³ and had a distinctive dentition consisting of procumbent incisors, blade-like premolars, molars with longitudinal rows of cusps (Fig. 1) and a predominantly posteriorly directed (palinal) chewing motion^{14,15}.

Palaeontologists have agreed for a long time that the success of multituberculate mammals was at least partly related to their highly derived dentition. Despite this, there is little consensus on the interpretations of their feeding ecology, perhaps owing to the limitations of previous approaches. For example, toothwear analysis is time intensive

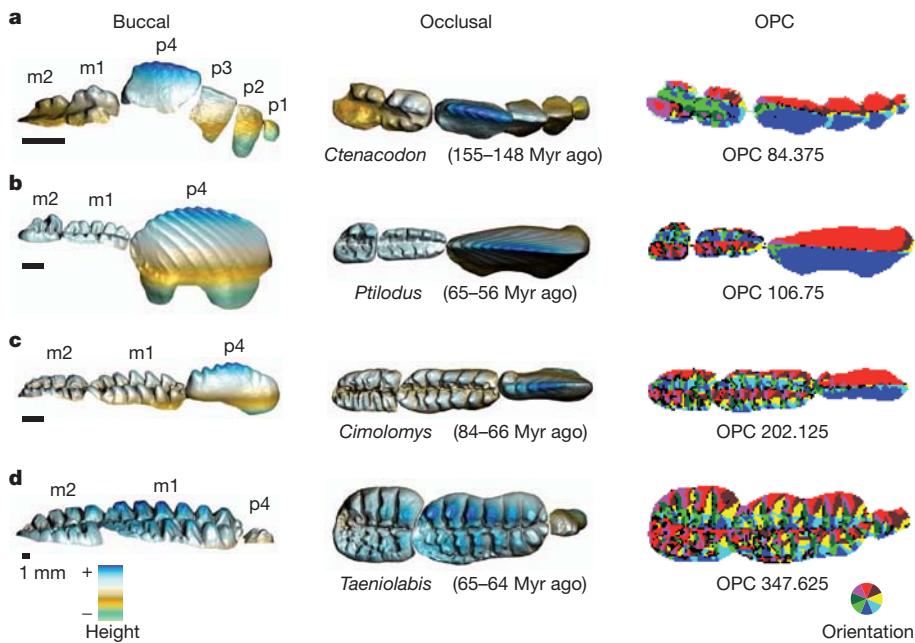


Figure 1 |

Dental and dietary diversity in multituberculate mammals. **a–d**, Three-dimensional buccal–occlusal and occlusal reconstructions of multituberculate lower-right cheek tooth rows for GIS analysis: Late Jurassic plagioulacid *Ctenacodon serratus* (**a**), Paleocene ptilodontoid *Ptilodus kummae* (**b**), Late

Cretaceous cimolomysid *Cimolomys gracilis* (**c**), Paleocene taeniolabidid *Taeniolabis taoensis* (**d**). Surface orientation map of each three-dimensional reconstruction (colour wheel indicates orientation) for OPC measurements (shown by the number of coloured patches). Clumps that are smaller than three grid points (black) are ignored. p, premolar; m, molar. Scale bars, 1 mm.

¹Department of Biology, University of Washington, Seattle, Washington 98195-1800, USA. ²School of Biological Sciences, Monash University, Victoria 3800, Australia. ³Developmental Biology Program, Institute of Biotechnology, University of Helsinki, PO Box 56, FIN-00014, Helsinki, Finland. ⁴Department of Geosciences and Geography, University of Helsinki, PO Box 64, FIN-00014, Helsinki, Finland.

and requires high-quality preservation¹⁵, and multituberculates lack living descendants and this hampers comparative studies. Interpretations of multituberculate feeding ecology therefore vary widely; they have been proposed to be broad herbivores, frugivores, granivores, root- and bark-eaters, egg-eaters, insectivores, carnivores and omnivores^{15–17}.

To obtain a robust and comprehensive view of multituberculate ecomorphological diversity through time, we quantified dental complexity in 41 genera using geographic information systems (GIS) analyses¹⁸ of three-dimensional crown surfaces of lower cheek teeth (Fig. 1; Supplementary Table 1). These analyses do not require cusp and facet homologies to be established, which can be a challenging task

when comparing morphologically and phylogenetically divergent taxa. Orientation patch count (OPC), a measure of dental complexity, was calculated as the number of discrete surfaces on the cheek tooth row distinguished by differences in orientation (for example, north, southwest; Fig. 1). Extant rodents, carnivorans and bats demonstrate a robust correlation between OPC and feeding ecology; OPC increases across the dietary spectrum from carnivores to omnivores to herbivores, despite many differences in specific tooth components, body size and chewing mechanics among these taxa^{18,19}. For a given clade, the standard deviation of OPC is an effective proxy for the dietary diversification and divergence in feeding function: higher standard deviation means greater dental disparity and a broader range of diets. OPC thus offers promise as a powerful tool for quantifying overall tooth shape and inferring diet in extinct taxa, such as multituberculates, that have highly derived dentitions with uncertain homology with living mammals and imprecise functional analogy.

Among the earliest multituberculates, the ‘Plagiaulacidae’ are a paraphyletic assemblage of taxa with up to four simple blade-like premolars and two multi-cusped molars (Fig. 1a). OPC analyses of ‘plagiaulacids’ from the Late Jurassic through to the Early Cretaceous epoch (from 156–100 Myr ago) show low and tightly constrained dental complexity (Fig. 2a; OPC, 84–125). Their OPC values correspond to carnivory and the low end of animal-dominated omnivory among extant mammals (for example, eating both insects and fruits). Multituberculates in the early Late Cretaceous (100–84 Myr ago), which include mostly basal members of the suborder Cimolodonta, had a slightly higher mean OPC than did the ‘plagiaulacids’ but retained the low standard deviation of OPC (Fig. 2a), indicating low morphological disparity. A distinct break occurred 84–66 Myr ago, in the latest Cretaceous, when the mean OPC rose and peaked within the Campanian (mean OPC, 145) and maximum OPC and disparity sharply increased as well (Fig. 2a; OPC, 70–230). Of the 17 taxa for this interval, 5 have OPC values that are greater than 160 and 2 have OPC values that are greater than 200, corresponding to values for plant-dominated omnivory and herbivory among extant mammals, respectively (Fig. 2a).

Finally, in the early Paleocene (66–62 Myr ago) multituberculates maintained high OPC (mean OPC, 138) and disparity peaked (Fig. 2a; OPC, 70–348). The early Paleocene *Taeniolabis* from North America (Fig. 1d) had the highest OPC among multituberculates (OPC, 348), which exceeded OPC values of extant herbivorous rodents and carnivorans¹⁸. Of the 16 other early Paleocene taxa, 4 have OPC values greater than 160 and 1 has an OPC value greater than 200 (Fig. 2a). Mean OPC and disparity of OPC declined during the remainder of the Paleocene and in the early Eocene (62–49 Myr ago). *Ectypodus*, the only known genus from the middle to late Eocene (49–35 Myr ago), has a low OPC (109), corresponding to the high end of the range for extant mammalian carnivores. This sharp drop in dental ecomorphological

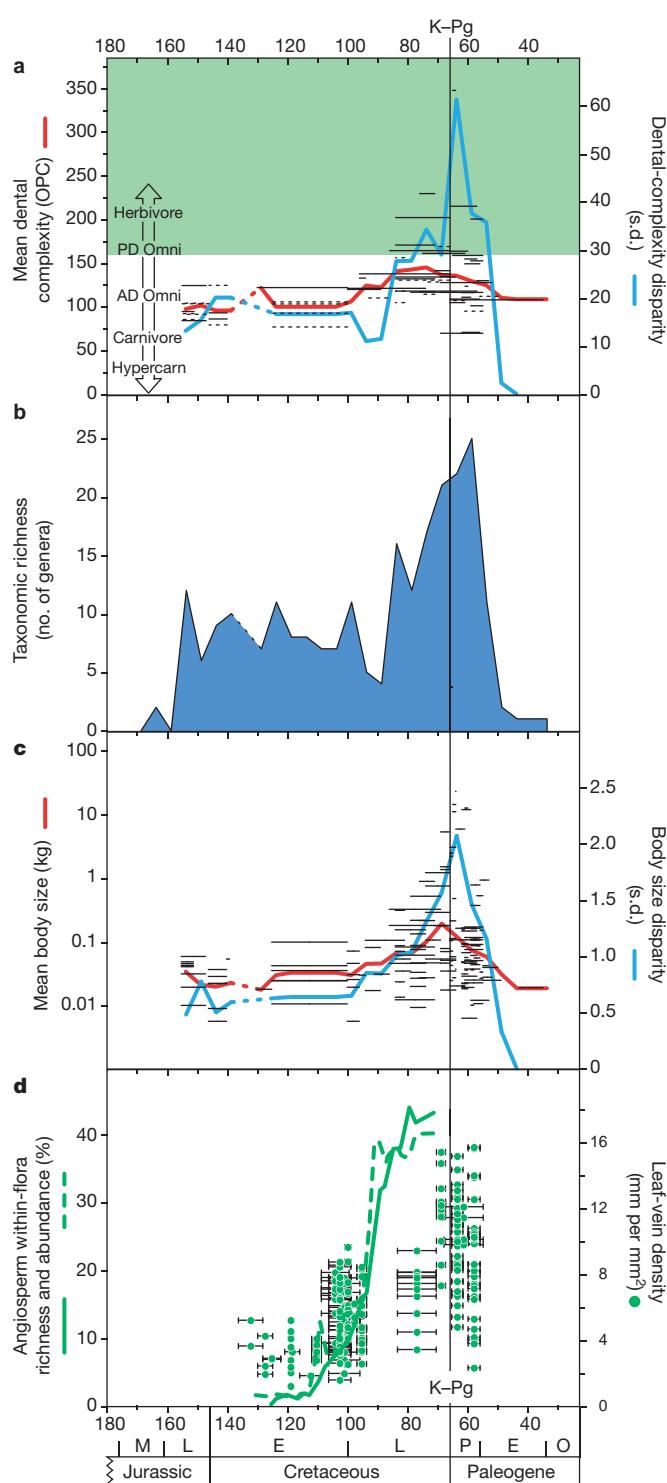


Figure 2 | Temporal patterns of multituberculate dental complexity, taxonomic richness, body size and angiosperm ecological diversification. a, Dental complexity (measured by OPC) for 41 multituberculate genera (solid black lines) and estimated for 24 additional genera (dashed black lines; see Supplementary Information) with mean dental complexity (OPC; red line) and disparity as standard deviations of OPC (blue line) in 5-Myr bins. Lengths of horizontal lines represent temporal ranges of taxa or uncertainties in ages of fossil localities. Labels for dietary classes are positioned at the lower end of their range based on OPC values in modern mammals¹⁸. Green shaded area represents plant-dominated omnivory (PD Omni) and herbivory. b, Taxonomic richness is equal to the number of genera per 5-Myr bin (blue shaded area). c, Body-mass estimates for 156 multituberculate species (solid black lines) with geometric means (red line) and disparity (blue line) for each 5-Myr bin. d, Angiosperm within-flora richness (solid green line) and relative abundance (dashed green line) from ref. 24 (± 7.5 -Myr moving averages). Leaf hydraulic capacity of angiosperms as leaf-vein density (green circles) from ref. 25. The 136–131-Myr bin was excluded from all analyses (Supplementary Information). AD Omni, animal-dominated omnivory; E, early; Hypercarn, hypercarnivory; M, middle; L, late; O, Oligocene; P, Paleocene.

disparity is intriguing in light of the hypothesis that some late Paleocene to early Eocene eutherian lineages (for example, rodents) competitively displaced multituberculates¹².

Variable sampling of fossils through time may influence our inferred patterns of multituberculate radiation. In particular, the relatively depauperate Jurassic and Early Cretaceous taxonomic richness of multituberculates may be partly due to the limited fossil record³. To test whether the changes in OPC through time might be sensitive to uneven sampling of fossils, we randomized the OPC value assignments for each genus. The results of 5,000 randomized mean OPC profiles through time indicate that even with additional fossil discoveries, the OPC patterns and ecological inferences that are presented here are likely to be robust (see Supplementary Fig. 8).

The overall trend of increasing multituberculate dental complexity was driven by increases in the number of cusps per molar and the relative size of molars in the cheek tooth row (Supplementary Fig. 1). This evolutionary pathway differs from that taken by ungulates, primates, rodents and lagomorphs, which responded to increased mechanical processing demands (that is, greater plant component) by evolving multi-cusped, molar-like premolars in addition to complex molars²⁰. Among several other mammalian lineages with blade-like premolars (for example, carbolestid primates and potoroid marsupials), only sthenurine marsupials re-evolved multi-cusped premolars²¹. Developmental, functional or structural constraints that are associated with the evolution of a blade-like premolar may have inhibited later evolution of more cuspidate premolars.

Patterns of generic richness and conservative estimates of body size for multituberculates (see Methods Summary) mirror the pattern of initially low dental complexity for most of the Mesozoic followed by a marked increase just before the end of the Cretaceous (84–66 Myr ago; Fig. 2b, c). From the Late Jurassic to the early Late Cretaceous, multituberculate generic richness fluctuated between 4 and 16 genera, and body mass ranged between 9 g and 105 g. A marked shift occurred near the end of the Cretaceous, with generic richness increasing from 16 to 21 genera and average body mass increasing from 67 g to 194 g, in particular, *Bubodens magnus* reached an estimated 5.25 kg (for a comparison, the alpine marmot weighs 3–8 kg). Similarly, Cenozoic generic richness and body size patterns parallel changes in dental complexity (Fig. 2b, c). However, it should be noted that body size and dental complexity are not interchangeable as predictors of diet because small multituberculates can have high OPC values, and large multituberculates can have low OPC values (Supplementary Information), conforming to previous analyses of extant taxa¹⁸. Furthermore, variation in the amount of available fossil-bearing rock through the studied interval could contribute to the pattern of generic richness²².

It is worth noting that increases in dental complexity (OPC of greater than 160) occurred in five multituberculate lineages: in the Asian Djadochtatherioidea, the North American Eucosmodontidae and Cimolomyidae (just before the end of the Cretaceous), the North American and Asian Taeniolabididae, and the North American Microcosmodontidae (in the early Paleocene). In the absence of a robust multituberculate phylogeny, it is unclear whether this represents a single evolutionary increase in tooth complexity, with multiple reversals to low complexity, or parallel increases in individual lineages or clades. This increase differs from the more commonly observed pattern, in which increased morphologic disparity precedes taxonomic diversification²³, as this increase occurred in step with an increase in multituberculate body size and taxonomic richness.

Overall, the pattern of increasing dental complexity in multituberculate mammals that pre-dates the Cretaceous–Paleogene mass extinction event contrasts with conventional ideas that mammalian evolution was suppressed during the Mesozoic era by selective pressures imposed by dinosaurs. Instead, multituberculates, the mammalian clade that co-existed for the longest time with non-avian dinosaurs, initiated an evolutionary radiation during the acme of dinosaur diversity, approximately 20 Myr before the Cretaceous–Paleogene boundary²².

This is consistent with the highly specialized adaptations that are found among several new, exceptionally well-preserved Mesozoic mammal specimens⁵ and with the timing of increased molecular divergence rates among extant mammalian lineages^{6,7}, but the multituberculates also show broad taxonomic and ecomorphologic diversification at the intraordinal level.

This adaptive response post-dated the taxonomic radiation of angiosperms but coincided broadly with increases in ecological diversity, abundance and leaf hydraulic capacities of angiosperms in the Late Cretaceous (in the Campanian and Maastrichtian)^{8,24,25} (Fig. 2d), suggesting that there is a causal link. Many angiosperms during this time were herbaceous, and had a rapid life cycle and less-effective herbivore defences compared to other seed plants and would consequently have been an attractive, protein-rich food source for herbivores²⁶. Some angiosperms had begun inhabiting a broader range of niches (trees, herbs and epiphytes) than most other plant groups and this may have enabled greater partitioning of the herbivore niche. Therefore, the multituberculate adaptive radiation, suggested by the increasing dental complexity, may have been triggered mainly by new niche space that was generated through the evolutionary and ecological radiations of angiosperms, but may have also been influenced by parallel radiations of associated non-angiosperm clades (for example, ferns²⁷), insect pollinators, dispersers and herbivores²⁸. A trophic link between angiosperms and some multituberculates is supported by our inferred dietary trend towards increased plant-dominated omnivory and herbivory among multituberculates just before the end of the Cretaceous and in the early Paleocene.

Our data also show that the dietary range of multituberculates did not decrease in response to the Cretaceous–Paleogene mass extinction event (Fig. 2a). Despite substantial taxonomic turnover of multituberculates at the Cretaceous–Paleogene boundary²⁹, they seem to have experienced little change in available food resources during one of the most severe extinction events in Earth's history. This apparent indifference of the multituberculate radiation to the Cretaceous–Paleogene event underscores the ecological selectivity of extinctions and suggests that broad-scale ecomorphological inventories of radiations may help to constrain the possible causes of extinction for other groups at the Cretaceous–Paleogene boundary^{9,10}.

METHODS SUMMARY

We scanned 48 dentitions from 41 genera of multituberculates. The three-dimensional scans of lower cheek tooth rows were analysed using GIS to quantify the number of discrete orientation patches using eight orientation directions. The mean of repeated measurements for eight rotations at multiples of 5.625° was used to reduce the effect of slight variations in the orientation of teeth. We compiled generic richness data from recent compendia and the primary literature. We estimated body mass from skull length, whenever available, and a tooth size to skull length regression formula. Detailed methods and calculations of dental complexity, generic richness and estimated body mass, data sets, measurements, calculations, randomization analyses and additional references are provided in the Supplementary Information.

Received 26 September 2011; accepted 20 January 2012.

Published online 14 March 2012.

1. Alroy, J. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* **48**, 107–118 (1999).
2. Smith, F. A. *et al.* The evolution of maximum body size of terrestrial mammals. *Science* **330**, 1216–1219 (2010).
3. Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
4. Van Valen, L. M. & Sloan, R. E. Ecology and the extinction of the dinosaurs. *Evol. Theory* **2**, 37–64 (1977).
5. Luo, Z.-X. Transformation and diversification in early mammalian evolution. *Nature* **450**, 1011–1019 (2007).
6. Bininda-Emonds, O. R. P. *et al.* The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007).
7. Meredith, R. W. *et al.* Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
8. Wing, S. L. & Boucher, L. D. Ecological aspects of the Cretaceous flowering plant radiation. *Annu. Rev. Earth Planet. Sci.* **26**, 379–421 (1998).

9. Archibald, J. D. *et al.* Cretaceous extinctions: multiple causes. *Science* **328**, 973 (2010).
10. Schulte, P. *et al.* The Chicxulub asteroid impact and mass extinction at the Cretaceous-Paleogene boundary. *Science* **327**, 1214–1218 (2010).
11. Weil, A. & Krause, D. W. in *Evolution of Tertiary Mammals of North America* Vol. 2 (eds Janis, C. M., Gunnell, G.F. & Uhen, M. D.) 19–38 (Cambridge Univ. Press, 2008).
12. Krause, D. W. in *Vertebrates, Phylogeny, and Philosophy* (eds Flanagan, K.M. & Lillegraven, J.A.) 119–130 (Contributions to Geology, 1986).
13. Rich, T. H. *et al.* An Australian multituberculate and its palaeobiogeographic implications. *Acta Palaeontol. Pol.* **54**, 1–6 (2009).
14. Gingerich, P. D. in *Patterns of Evolution* (ed. Hallam, A.) 469–500 (Elsevier, 1977).
15. Krause, D. W. Jaw movement, dental function, and diet in the Paleocene multituberculate *Ptilodus*. *Paleobiology* **8**, 265–281 (1982).
16. Cope, E. D. The tertiary Marsupialia. *Am. Nat.* **18**, 686–697 (1884).
17. Simpson, G. G. The “plagiaulacoid” type of mammalian dentition. *J. Mamm.* **14**, 97–107 (1933).
18. Evans, A. R., Wilson, G. P., Fortelius, M. & Jernvall, J. High-level similarity of dentitions in carnivorans and rodents. *Nature* **445**, 78–81 (2007).
19. Santana, S. E., Strait, S. & Dumont, E. R. The better to eat you with: functional correlates of tooth structure in bats. *Funct. Ecol.* **25**, 839–847 (2011).
20. Jernvall, J., Gilbert, C. C. & Wright, P. C. in *Elwyn Simons: A Search for Origins* (eds Fleagle, J. G. & Gilbert, C. C.) 335–342 (Springer, 2008).
21. Prideaux, G. J. Systematics and evolution of the sthenurine kangaroos. *Univ. Calif. Publ. Geol. Sci.* **146**, 1–622 (2004).
22. Barrett, P. M., McGowan, A. J. & Page, V. Dinosaur diversity and the rock record. *Proc. R. Soc. B* **276**, 2667–2674 (2009).
23. Foote, M. The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.* **28**, 129–152 (1997).
24. Lupia, R., Lidgard, S. & Crane, P. R. Comparing palynological abundance and diversity: implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology* **25**, 305–340 (1999).
25. Feild, T. S. *et al.* Fossil evidence for Cretaceous escalation in angiosperm leaf vein evolution. *Proc. Natl Acad. Sci. USA* **108**, 8363–8366 (2011).
26. Wing, S. L. & Tiffney, B. H. in *The Origins of Angiosperms and Their Biological Consequences* (eds Friis, E.M., Chaloner, W.G. & Crane, P.R.) 203–224 (Cambridge Univ. Press, 1987).
27. Schneider, H. *et al.* Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553–557 (2004).
28. Grimaldi, D. The co-radiations of pollinating insects and angiosperms in the Cretaceous. *Ann. Mo. Bot. Gard.* **86**, 373–406 (1999).
29. Wilson, G. P. Mammalian faunal dynamics during the last 1.8 million years of the Cretaceous in Garfield County, Montana. *J. Mamm. Evol.* **12**, 53–76 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature/.

Acknowledgements We thank museums, institutions and individuals that made specimens available for this study (full list is available in Supplementary Information). Funding was provided by the National Science Foundation, Denver Museum, the University of Washington (G.P.W. and P.D.S.), the Australian Research Council, Monash University (A.R.E.), the Academy of Finland (A.R.E., M.F. and J.J.) and the EU SYNTHESYS program (project GB-TAF-4779) (I.J.C.).

Author Contributions G.P.W., A.R.E., J.J. and M.F. designed the study. G.P.W., A.R.E., I.J.C. and P.D.S. collected and analysed the data. G.P.W., A.R.E. and J.J. wrote the manuscript. G.P.W., A.R.E., I.J.C., P.D.S., M.F. and J.J. discussed results and commented on the manuscript at all stages.

Author Information The three-dimensional scans for this study are deposited in the MorphoBrowser database (<http://morphobrowser.biocenter.helsinki.fi/>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.P.W. (gwilson@u.washington.edu).

Identification and characterization of a bacterial hydrosulphide ion channel

Bryan K. Czyzewski^{1,2} & Da-Neng Wang^{1,3}

The hydrosulphide ion (HS^-) and its undissociated form, hydrogen sulphide (H_2S), which are believed to have been critical to the origin of life on Earth¹, remain important in physiology and cellular signalling². As a major metabolite in anaerobic bacterial growth, hydrogen sulphide is a product of both assimilatory and dissimilatory sulphate reduction^{2–4}. These pathways can reduce various oxidized sulphur compounds including sulphate, sulphite and thiosulphate. The dissimilatory sulphate reduction pathway uses this molecule as the terminal electron acceptor for anaerobic respiration, in which process it produces excess amounts of H_2S (ref. 4). The reduction of sulphite is a key intermediate step in all sulphate reduction pathways. In *Clostridium* and *Salmonella*, an inducible sulphite reductase is directly linked to the regeneration of NAD^+ , which has been suggested to have a role in energy production and growth, as well as in the detoxification of sulphite⁵. Above a certain concentration threshold, both H_2S and HS^- inhibit cell growth by binding the metal centres of enzymes and cytochrome oxidase⁵, necessitating a release mechanism for the export of this toxic metabolite from the cell^{5–9}. Here we report the identification of a hydrosulphide ion channel in the pathogen *Clostridium difficile* through a combination of genetic, biochemical and functional approaches. The HS^- channel is a member of the formate/nitrite transport family, in which about 50 hydrosulphide ion channels form a third subfamily alongside those for formate^{10,11} (FocA) and for nitrite¹² (NirC). The hydrosulphide ion channel is permeable to formate and nitrite as well as to HS^- ions. Such polyspecificity can be explained by the conserved ion selectivity filter observed in the channel's crystal structure. The channel has a low open probability and is tightly regulated, to avoid decoupling of the membrane proton gradient.

Because H_2S is a weak acid with a $\text{p}K_a$ of 6.8 and a second $\text{p}K_a$ of 19, in a biological system more than 80% of it will exist in its ionized form as HS^- (refs 2, 13). To account for membrane permeation of intracellularly produced H_2S (refs 6, 9), a channel for HS^- or H_2S has been suggested^{5,7}, with the water channel aquaporins as possible candidates⁸. Whereas studies in planar lipid bilayers indicate that the aquaporin from *Archaeoglobus fulgidus* is not permeable to H_2S and that the lipid bilayer provides little resistance to H_2S permeation⁹, in the vent worm *Riftia pachyptila* it has been shown that HS^- ions, but not H_2S , are selectively transported through its outer epithelium into its vasculature⁵.

Despite the high permeability of H_2S across the lipid bilayer, we argue that its weak acidity necessitates a release mechanism for its ionized form, HS^- (Fig. 1a). As a weak acid with a $\text{p}K_a$ of 6.8, H_2S can diffuse readily across the lipid bilayer. Therefore, the distribution of H_2S and HS^- on the two sides of the cell membrane is directly proportional to the pH differential as described by the Henderson–Hasselbalch equation¹⁴. Because cellular pH is kept at neutral levels, and extracellular pH is typically one to two units more acidic, at equilibrium the HS^- concentration will be 10–100-fold greater on the inside of the cell than on the outside in the absence of a release

mechanism for the HS^- anions. Therefore, given its significant toxicity, it would be beneficial to directly expel this ion across the membrane by the quickest mechanism available: an ion channel.

Recently, a novel family of anion channels for short-chain acids has been functionally characterized^{11,15,16}. Members of the formate/nitrite transport (FNT) family transport various anions during anaerobic bacterial growth^{10,12,17}. Judging from sequence homology, however, only about half of its 2,000 identified members clearly belong to either the FocA or the NirC subfamily (Figs 1b, c and Supplementary Figs 1 and 2). Notably, for both FocA and NirC, each channel gene is genetically linked to its reductase partner: *focA* to the formate reductase *pflB*¹⁰ and *nirC* to the nitrate reductase *nirBD*¹² (Fig. 1b). Our phylogenetic analysis of bacterial genes suggests the existence of additional

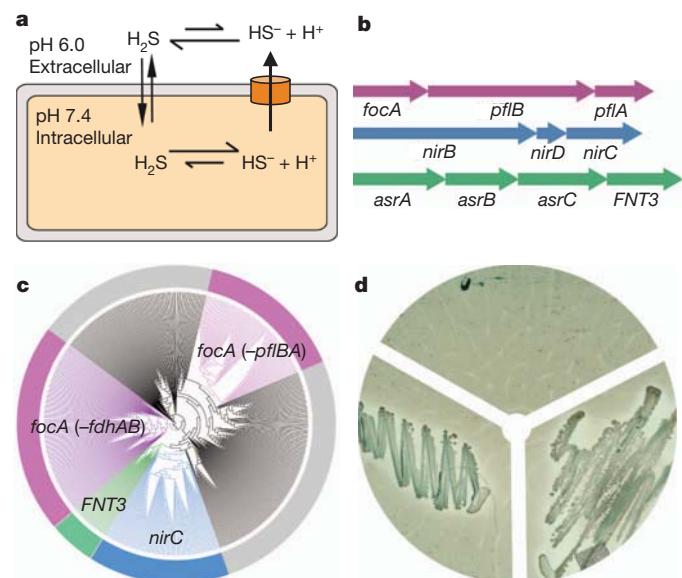


Figure 1 | Genetic analyses and functional characterization of FNT3 and the *asrABC* operon. About 50 FNT genes from *Clostridium* are linked to the sulphite reductase gene *asrABC*, and they form a third subfamily (FNT3) in the FNT family. **a**, Model of the intracellular anion concentrative effect for the weak acid H_2S . **b**, Genomic organization of three types of FNT gene and their metabolically related reductase genes. Representatives are shown for *focA*, *nirC* and *FNT3* with their respective linked operons. **c**, Phylogenetic tree of 474 bacterial and archaeal members of the FNT family. Branches are coloured on the basis of genetic linkage to metabolic enzymes: genes linked to pyruvate formate lyase (*pflBA*) or formate dehydrogenase (*fdhAB*; *fdhA* and *fdhB* also known as *selA* and *selD*, respectively) are coloured pink, genes linked to nitrite reductase (*nirBD*) are coloured blue and genes linked to sulphite reductase (*asrABC*) are coloured green. The FocA protein in archaea is encoded by *fdhC* (*selC*). The grey areas represent FNT family members with no assigned function based on genetic linkage. **d**, Bismuth sulphite agar plate assay: vector control (top); *asrA*, *asrB* and *asrC* (left); *asrA*, *asrB*, *asrC* and *FNT3* (right).

¹The Helen L. and Martin S. Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA. ²Molecular Biophysics Graduate Program, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA. ³Department of Cell Biology, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA.

channel subfamilies in the FNT family (Fig. 1c and Supplementary Fig. 1). One such subfamily is a group of ~ 50 homologous genes, temporarily termed FNT3, found in various species of *Clostridium* that are major human pathogens grown under strictly anaerobic conditions. As it is linked to a reductase gene, the *asrABC* operon¹⁸ (Fig. 1b), we proposed that FNT3 genes might encode a channel for HS^- or related ions. *AsrABC* reduces sulphite (SO_3^{2-}) to sulphide (S^{2-}), which is rapidly converted into H_2S and HS^- under physiological conditions.

Salmonella asrABC is the only such operon that has been characterized thus far. Its overexpression in *Escherichia coli*, which lacks this biochemical pathway, was previously shown to reduce SO_3^{2-} to H_2S (ref. 19). To verify first whether the homologous *asrABC* operon from *C. difficile* also encodes an SO_3^{2-} reductase, we tested the growth of *E. coli* transformed with the *C. difficile asrABC* genes, plated on bismuth sulphite agar^{19,20}. When *C. difficile asrABC* was induced, *E. coli* was able to overcome the expected growth inhibition by reducing SO_3^{2-} to H_2S (Fig. 1d and Supplementary Fig. 3). Darkening along the edges of the colonies, formed by the precipitation of bismuth and iron sulphide, further indicated that H_2S gas was being produced. This showed that the *asrABC* genes from *C. difficile* do indeed encode for an SO_3^{2-} reductase, which in turn supports the notion that the FNT3 gene ('FNT3') in the *asrABC* operon may encode a channel for SO_3^{2-} or for its reduced product, HS^- .

To determine whether the FNT3 protein from *C. difficile* functions as a channel for HS^- or SO_3^{2-} , we then carried out whole-cell transport, ion–protein binding in solution and transport-inhibition assays in reconstituted proteoliposomes. First we transformed sulphide-producing *Salmonella typhimurium*, which lacks an HS^- channel, with the *C. difficile FNT3*. In SO_3^{2-} -supplemented minimal medium, the *FNT3*-expressing *S. typhimurium* released a high concentration of HS^- ions into the culture medium (Fig. 2a and Supplementary Fig. 4). This increase in extracellular HS^- was specific to HS^- produced by the endogenous cytoplasmic *AsrABC*; when the other sulphide-producing reductase in the cell, the periplasmic thiosulphate reductase, was engaged with its substrate thiosulphate, little effect on the extracellular concentration of HS^- was observed. This indicates that the measured HS^- was generated by SO_3^{2-} reduction by the cytoplasmic *AsrABC* and then exported. Such an FNT3-linked increase in HS^- production in *S. typhimurium* can be attributed to either increased import of SO_3^{2-} or increased export of HS^- from the cell. It follows that an FNT3 protein could be an ion channel for SO_3^{2-} or HS^- .

An ion channel protein is often stabilized by its permeating ions through their direct interaction with its selectivity filter²¹. Therefore, we tested whether purified FNT3 protein interacts with the monovalent HS^- or the divalent SO_3^{2-} . Using size-exclusion chromatography of purified protein samples incubated at different temperatures, we found that at pH 8.0 the presence of HS^- ions was able to increase the nominal melting temperature of the FNT3 protein by 8 °C, whereas SO_3^{2-} had little effect (Fig. 2b and Supplementary Table 1). This result is consistent with the hypothesis that FNT3 protein functions as an HS^- channel but is probably impermeable to SO_3^{2-} . Notably, three other monovalent anions, formate, nitrite and chloride, also increased the melting temperature of the FNT3 protein, but to a lesser degree (3–4 °C).

We went on to identify the permeable ions of the FNT3 protein *in vitro*. As channels in the FNT family are expected to have low conductance rates¹¹, we measured the transport activity of FNT3 protein using a concentrative uptake assay^{11,22}. Given that HS^- ions at concentrations needed to set up a sufficient electrochemical gradient (~ 150 mM) are severely disruptive to the membrane, we chose to measure the permeability of FNT3 protein using another interacting ion, formate (Fig. 2c). Proteoliposomes reconstituted at pH 8.0 from purified *C. difficile* FNT3 protein were found to be permeable to formate, although at a lower rate than FocA from *Vibrio cholerae* (Fig. 2c). This permeability to formate by FNT3 protein allowed us to search for other ions that directly compete with formate for transport. Indeed, when HS^- or nitrite was present in the buffer outside the proteoliposomes,

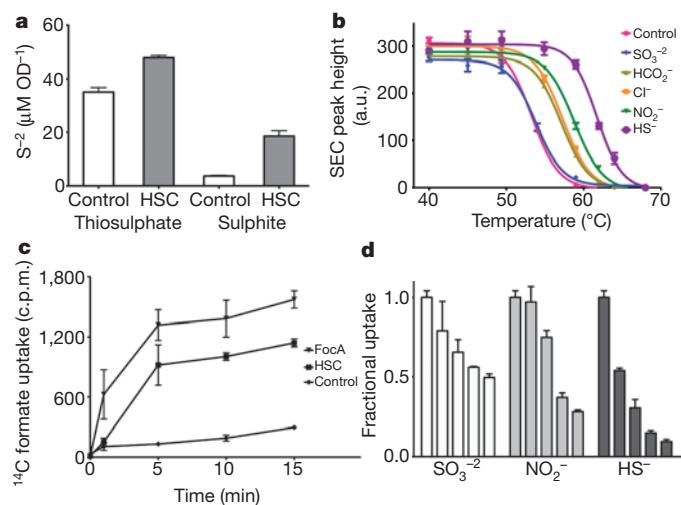


Figure 2 | Binding and transport activity of HSC in reconstituted proteoliposomes. **a**, Measurements of sulphide concentrations in the media of *Salmonella* transformed with vector control or vector encoding HSC. Minimal media were supplemented with either sulphite or thiosulphate to induce the production of hydrogen sulphide from either periplasmic thiosulphate reductase or cytoplasmic sulphite reductase. OD, optical density. **b**, Binding of detergent-solubilized and purified HSC to various anions was determined by using thermostability coupled size-exclusion chromatography. Peak heights of recovered samples are plotted against temperature and fitted to a Boltzmann sigmoidal model to determine nominal melting temperatures. **c**, Radiolabelled formate uptake in proteoliposomes reconstituted with purified HSC at pH 8.0 was monitored in a concentrative uptake assay and compared with FocA activity or vesicle controls. c.p.m., counts per minute. **d**, Inhibition of radiolabelled concentrative uptake of formate by the addition of various anions at increasing concentrations. The graph shows the amount of radiolabelled formate measured at the 10-min time point for each concentration of anion tested. The concentrations of the competing anions were 0, 0.15, 0.6, 3 and 15 mM (left to right). Error bars, s.e.m. ($N = 3$).

the uptake of formate was drastically inhibited (Fig. 2d). In the physiological sulphide concentration range (1–2 mM), $\sim 70\%$ inhibition was observed, whereas at the highest concentration tested (15 mM), HS^- and nitrite inhibited the uptake of formate by 90% and 70%, respectively.

Such direct competitive inhibition (Fig. 2c), coupled with whole-cell transport assays (Fig. 2a) and *in vitro* binding assays (Fig. 2b), collectively indicates that the FNT3 protein from *C. difficile* is likely to be a channel for the hydrosulphide ion, HS^- . Although the channel is also permeable to formate and nitrite ions, the FNT3 gene is directly linked to the *asrABC* sulphite reductase operon (Fig. 1b); we have therefore tentatively named the FNT3 protein a hydrosulphide ion channel (HSC). A definite understanding of the physiological role of these channels will need to be confirmed by *in vivo* experiments such as genetic deletion studies.

To understand the transport mechanism of the HSC protein from *C. difficile* (Supplementary Fig. 5), we determined its crystal structure at high pH (pH 9.0) to a resolution of 2.2 Å (Fig. 3a and Supplementary Table 2). The HSC adopts the aquaporin/FocA fold^{11,23} and forms a pentamer like FocA, with the five protomers having the same structure (Supplementary Figs 6–8). The two-fold inverted symmetry between the two halves of the protein is greater than that observed in FocA and extends to the pore-lining transmembrane helices TM2b and TM5b (Supplementary Fig. 8b), two salt-bridge triads on the periplasmic and the cytoplasmic sides (Lys 148–Glu 200–Asn 205 and Lys 16–Glu 81–Asn 86) (Fig. 4a–c), and the two short helices parallel to the membrane (helix P on the periplasmic surface and helix N at the amino terminus on the cytoplasmic surface) (Supplementary Fig. 8b). This helix N, by means of the Lys 16–Glu 81–Asn 86 salt-bridge triad, buttresses TM2b and TM5a, making the cytoplasmic side of the protein more closed and rigid than that in FocA¹¹.

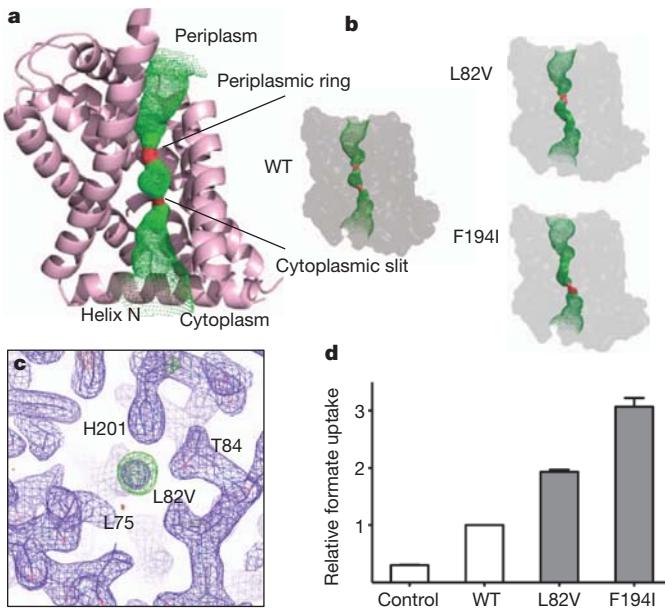


Figure 3 | Structural and functional characterization of the ion permeation pathway. **a**, Structure of a protomer of HSC overlaid with the pore diameter calculations made using HOLE, coloured to indicate permeability to water: green, permeable; red, impermeable. Transmembrane helix 2 has been removed for clarity. **b**, Pore diameter calculations from HOLE of the crystal structures for each of the two permeation pathway mutations, Leu 82 Val and Phe 194 Ile and wild type (WT). **c**, Close-up of the electron density observed at the selectivity filter for the Leu 82 Val mutant. The $2F_o - F_c$ electron density map contoured at 1.1σ is coloured purple and the $F_o - F_c$ difference density map contoured at 3σ is coloured green. **d**, Relative uptake of proteoliposomes reconstituted with purified HSC and pathway mutations. The graph shows the 10-min time point of each concentrative uptake experiment. Error bars, s.e.m. ($N = 3$).

The pore for ion permeation, which is located at the centre of each protomer, contains the selectivity filter formed by a cytoplasmic slit and a periplasmic ring in the middle of the membrane (Fig. 3a, b). The radius of the cytoplasmic slit is 0.8 Å, which is too small for an HS^- ion (radius, 1.7 Å; ref. 24) to pass through (Supplementary Figs 9–11).

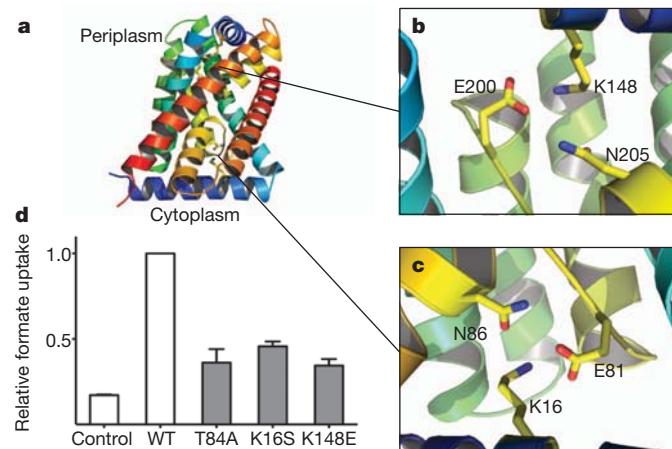


Figure 4 | Structural and functional characterization of possible gating mechanisms. **a**, Structure of the HSC protomer representing possible gating regions. The structure is coloured to show the two-fold inverted topology. **b, c**, The Glu-Lys-Asn salt-bridge triads, related by a pseudo-two-fold symmetry, help to stabilize helix P at the periplasmic side (**b**) and helix N at the cytoplasmic side (**c**) of the protein. The rotamer that each residue adopts is conserved. **d**, Relative uptake of proteoliposomes reconstituted with purified HSC and gating mutations. The graph shows the 10-min time point of each concentrative uptake experiment. Error bars, s.e.m. ($N = 3$).

Therefore, the HSC protein in the crystal structure is probably in a closed state. As expected from the sequence similarity between HSC and FocA in the selectivity filter (Supplementary Fig. 2), the positions or orientations of key residues changed only slightly in HSC (Supplementary Fig. 10). Such conservation explains the permeability of HSC to HS^- , as well as to formate and nitrite. When a residue at the cytoplasmic slit or the periplasmic ring is mutated into a smaller residue, for example Leu 82 to Val or Phe 194 to Ile, the pore size is increased, as shown in the crystal structures of the mutant protein (Fig. 3b and Supplementary Table 2), and the formate transport rate of the channel is significantly increased (Fig. 3d). We note that in the crystal structure of the Leu 82 Val mutant, electron density was observed in the cytoplasmic slit (Fig. 3c). This density, which is greater than that of a water molecule—as inferred from a B-factor fourfold lower than those of the surrounding atoms—could be accounted for by a Cl^- or an HS^- ion. Owing to their similar sizes and chemical properties, it has been previously observed that Cl^- and HS^- ions often bind to the same site in protein crystal structures²⁵.

The observed polyspecificity of HSC is typical for anion channels^{26–28}, and is most probably shared by the FocA and NirC channels as well. Such polyspecific anion recognition is at least partly due to the diffuse nature of the electron cloud of anions²⁹. By contrast, cation channels often show a high degree of selectivity, which is made possible by the point-like charge properties of such ions. Because different FNT channels in this ancient protein family are coupled with their specific metabolic pathways, each channel is expected to have its own preference for permeating ions, and that preference is probably due to the small structural variations in the selectivity filter that we observed here (Supplementary Fig. 10).

The expulsion of toxic concentrations of HS^- from the cytoplasm through HSC needs to be balanced against the potential risk of the uncoupling of the membrane proton gradient. Freely diffusible H_2S and other weak, short-chain acids will dissociate into an anion and a proton in the cytoplasm (Fig. 1a). As the anion is expelled, there will be an intracellular net accumulation of protons. Therefore, the ground state of HSC is likely to be closed and its opening tightly regulated.

Indeed, the HSC structure we determined at high pH is in its closed state (Fig. 3 and Supplementary Figs 9–11). We then considered how easily the channel could be opened. Because the homologous FocA channel can be gated by an appropriate change in pH or cytoplasmic formate concentration^{11,16}, we first investigated possible pH-dependent structural changes of HSC by determining its crystal structure at pH 7.5 (resolution, 3.2 Å) and at pH 4.5 (3.0 Å) (Supplementary Table 2). The protein structures obtained at these neutral and low pH values were the same as the structure at 2.2 Å and pH 9.0 (Supplementary Fig. 12), and the channel remained closed. We then studied the two salt-bridge triads, both of which contain a glutamate residue that could potentially function as a pH sensor (Fig. 4a–c). When either of the salt bridges was broken, by mutating Lys 16 to Ser or Lys 148 to Glu, the HSC crystal structure did not change and the channel still remained closed (Supplementary Fig. 13 and Supplementary Table 2). Consistent with these structural results, the transport rates of these mutants did not increase (Fig. 4d).

To determine whether HSC is gated by anion concentration in a similar manner to *V. cholerae* FocA¹¹, in which the competition of formate with Thr 90 from the Ω -loop for hydrogen bonding to His 208 in the selectivity filter opens the channel, we mutated the equivalent threonine in HSC, Thr 84, to an Ala and determined its crystal structure at 2.4 Å (Supplementary Table 2). This mutant structure did not differ from the closed wild-type channel, and no increase in transport was detected (Fig. 4d and Supplementary Fig. 14). Because Thr 84, Lys 16 and Lys 148 all line the ion permeation pathway, mutation of these residues slightly reduced the ion permeation rate of the channel. Therefore, we conclude that the opening of HSC is indeed tightly regulated and the channel has a much lower open probability than FocA^{11,16}.

The low open probability of HSC observed here is consistent with its physiological role in the cell. The opening of HSC probably requires significant movement of TM2b. As TM2b is 'held' in space by the cytoplasmic helix N (Fig. 4a), the movement of the latter, for example by direct interaction with a cytoplasmic enzyme in the same metabolic pathway, can trigger the opening of the channel. We have no evidence for such a hypothesis, and it can be validated only by further experimentation, which we expect would reveal greater variations in both the gating mechanism and the anionic preference of FNT channels.

METHODS SUMMARY

Bismuth sulphite agar plate assay. *FNT3* was isolated from *C. difficile* strain 630. *Escherichia coli* BL21(DE3) transformed with vector pCDFduet-1-asrA-asrB, pACYCduet-1-asrC, or pACYCduet-1-asrC-FNT3 were plated on bismuth sulphite agar²⁰ and allowed to develop for 48 h.

Whole-cell *in vivo* transport. *Salmonella typhimurium* transformed with pBAD-FNT3 were assayed for their ability to generate sulphide from either thiosulphate or sulphite using a silver–sulphide probe following modification of a previously published protocol¹⁹.

Proteoliposome reconstitution and *in vitro* transport assay. Proteoliposomes were formed from *E. coli* total lipids and Ni²⁺-NTA-purified proteins and assayed by a method modified from a previously published concentrative uptake assay^{11,22}.

Protein–ion interaction assay. Ion-induced thermostability of HSC was monitored using analytical size-exclusion chromatography. Purified protein was incubated at increasing temperatures and its recovery was monitored by size-exclusion chromatography using a high-performance liquid chromatograph. Candidate permeating ions were incubated at a concentration of 20 mM and protein was monitored for thermostabilization.

Protein purification and crystallization and structure determination. HSC was expressed as a carboxy-terminal FLAG–10×His fusion in *E. coli* BL21. The protein was purified using Ni²⁺-NTA affinity chromatography, followed by size-exclusion chromatography. The protein was crystallized by hanging-drop vapour diffusion and the crystal structure was solved by molecular replacement.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 August 2011; accepted 20 January 2012.

Published online 11 March 2012.

- Wächtershäuser, G. Groundworks for an evolutionary biochemistry: the iron-sulphur world. *Prog. Biophys. Mol. Biol.* **58**, 85–201 (1992).
- Kabil, O. & Banerjee, R. Redox biochemistry of hydrogen sulfide. *J. Biol. Chem.* **285**, 21903–21907 (2010).
- Dhillon, A., Goswami, S., Riley, M., Teske, A. & Sogin, M. Domain evolution and functional diversification of sulfite reductases. *Astrobiology* **5**, 18–29 (2005).
- Rabus, R., Hansen, T. & Widdel, F. in *The Prokaryotes: Ecophysiology and Biochemistry* Vol. 2 (eds Dworkin, M. et al.) 659–768 (Springer, 2006).
- Goffredi, S. K., Childress, J. J., Desaulniers, N. T. & Lallier, F. J. Sulfide acquisition by the vent worm *Riftia pachyptila* appears to be via uptake of HS[−], rather than H₂S. *J. Exp. Biol.* **200**, 2609–2616 (1997).
- Jacques, A. G. The kinetics of penetration: XII. Hydrogen sulfide. *J. Gen. Physiol.* **19**, 397–418 (1936).
- Freitag, J. K. et al. A paradox resolved: sulfide acquisition by roots of seep tubeworms sustains net chemoautotrophy. *Proc. Natl Acad. Sci. USA* **98**, 13408–13413 (2001).
- Lee, J. K. et al. Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 Å. *Proc. Natl Acad. Sci. USA* **102**, 18932–18937 (2005).
- Mathai, J. C. et al. No facilitator required for membrane transport of hydrogen sulfide. *Proc. Natl Acad. Sci. USA* **106**, 16633–16638 (2009).
- Suppmann, B. & Sawers, G. Isolation and characterization of hypophosphite-resistant mutants of *Escherichia coli*: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. *Mol. Microbiol.* **11**, 965–982 (1994).
- Waight, A. B., Love, J. & Wang, D. N. Structure and mechanism of a pentameric formate channel. *Nature Struct. Mol. Biol.* **17**, 31–37 (2010).

- Jia, W., Tovell, N., Clegg, S., Trimmer, M. & Cole, J. A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *Biochem. J.* **417**, 297–304 (2009).
- Hughes, M. N., Centelles, M. N. & Moore, K. P. Making and working with hydrogen sulfide: the chemistry and generation of hydrogen sulfide *in vitro* and its measurement *in vivo*: a review. *Free Radic. Biol. Med.* **47**, 1346–1353 (2009).
- Hirshfield, I. N., Terzulli, S. & O’Byrne, C. Weak organic acids: a panoply of effects on bacteria. *Sci. Prog.* **86**, 245–269 (2003).
- Wang, Y. et al. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462**, 467–472 (2009).
- Lu, W. et al. pH-dependent gating in a FocA formate channel. *Science* **332**, 352–354 (2011).
- Das, P., Lahiri, A. & Chakravortty, D. Novel role of the nitrite transporter NirC in *Salmonella* pathogenesis: SPI2-dependent suppression of inducible nitric oxide synthase in activated macrophages. *Microbiology* **155**, 2476–2489 (2009).
- Crane, B. R. & Getzoff, E. D. The relationship between structure and function for the sulfite reductases. *Curr. Opin. Struct. Biol.* **6**, 744–756 (1996).
- Hallenbeck, P. C., Clark, M. A. & Barrett, E. L. Characterization of anaerobic sulfite reduction by *Salmonella typhimurium* and purification of the anaerobically induced sulfite reductase. *J. Bacteriol.* **171**, 3008–3015 (1989).
- Wilson, W. J. & Blair, E. M. M.v. A combination of bismuth and sodium sulphite affording an enrichment and selective medium for the typhoid-paratyphoid groups of bacteria. *J. Pathol. Bacteriol.* **29**, 310–311 (1926).
- Doyle, D. A. et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Middleton, R. E., Pheasant, D. J. & Miller, C. Purification, reconstitution, and subunit composition of a voltage-gated chloride channel from *Torpedo* electroplax. *Biochemistry* **33**, 13189–13198 (1994).
- Savage, D. F., O’Connell, J. D. III, Miercke, L. J., Finer-Moore, J. & Stroud, R. M. Structural context shapes the aquaporin selectivity filter. *Proc. Natl Acad. Sci. USA* **107**, 17164–17169 (2010).
- Feth, S., Gibbs, G. V., Boisen, M. B. Jr & Myers, R. H. Promolecule radii for nitrides, oxides, and sulfides. A comparison with effective ionic and crystal radii. *J. Phys. Chem.* **97**, 11445–11450 (1993).
- Tai, C. H. et al. Characterization of the allosteric anion-binding site of O-acetylserine sulfhydrylase. *Biochemistry* **40**, 7446–7452 (2001).
- Hille, B. *Ionic Channels of Excitable Membranes* 362–389 (Sinauer, 1992).
- Yasui, M. et al. Rapid gating and anion permeability of an intracellular aquaporin. *Nature* **402**, 184–187 (1999).
- Rychkov, G. Y., Pusch, M., Roberts, M. L., Jentsch, T. J. & Bretag, A. H. Permeation and block of the skeletal muscle chloride channel, ClC-1, by foreign anions. *J. Gen. Physiol.* **111**, 653–665 (1998).
- Simons, J. & Jordan, K. D. Ab initio electronic structure of anions. *Chem. Rev.* **87**, 535–555 (1987).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful the staff at beamlines X25 and X29 of the National Synchrotron Light Source at Brookhaven National Laboratory and at beamline 23ID of the Advanced Photon Source at Argonne National Laboratory for assistance in X-ray diffraction experiments. We thank A. B. Waight for suggesting the project; J. J. Marden for assistance with cloning of mutants; T. Neubert and S. Blais for mass spectrometry measurements; the 2010 CCP4 Workshop for assistance in processing diffraction data; and A. David, H. Jackson, N. K. Karpowich, J. J. Marden, R. L. Mancuso, Y. Pan and M. Zhou for discussions. This work was financially supported by the NIH (R01-GM093825, R01-DK073973, R01-MH083840 and U54-GM075026). B.K.C. was partly supported by an NIH Supplement Grant to Promote Diversity in Health-Related Research (R01-DK053973-08A1S1) and an NIH pre-doctoral fellowship (F31-AI086072).

Author Contributions B.K.C. did the experiments. B.K.C. and D.-N.W. wrote the manuscript.

Author Information The atomic coordinates and structure factors of HSC for high, medium and low pH have been deposited in the Protein Data Bank under accession codes 3TDO, 3TDR and 3TDP, respectively, and those of the Lys 16 Ser, Leu 82 Val, Thr 84 Ala, Lys 148 Glu and Phe 194 Ile mutants have been deposited under the codes 3TE2, 3TDX, 3TE1, 3TE0 and 3TDS, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.N.W. (wang@saturn.med.nyu.edu).

METHODS

Phylogenetic analysis. Phylogenetic data for 474 bacterial and archaeal members of the FNT family were obtained from the HOMOGENOM database³⁰, aligned using the program JALVIEW³¹ and plotted using the program ARCHAEOPTERYX³². Genomic sequences and operon annotations were derived from the NCBI GenBank nucleotide database³³.

Bismuth sulphite agar plate assay. Gene *FNT3* was isolated from *C. difficile* strain 630. The genes for *C. difficile* *asrA* (NCBI gene ID, 4915353) and *asrB* (4915352) were cloned into pCDFDuet-1 (EMD Biosciences) and the genes for *C. difficile* *asrC* (4915351) and *FNT3* (4915350) were cloned into pACYCDuet-1 (EMD Biosciences). Bismuth sulphite agar plates²⁰ were freshly prepared following the manufacturer's protocol and supplemented with 1 mM isopropyl β-D-1-thiogalactopyranoside and the antibiotics chloramphenicol and spectinomycin. *E. coli* BL21(DE3) cells were transformed with the indicated vectors and streaked, and colonies were allowed to grow at 37 °C for 48 h. The bismuth sulphite agar²⁰ contained the ATP synthase inhibitor brilliant green¹⁹ and high concentrations of SO₃⁻². When *C. difficile* *asrABC* was induced, *E. coli* were able to overcome the growth inhibition by producing ATP through glycolysis by reducing SO₃⁻² to H₂S.

Whole-cell *in vivo* transport assay. *Salmonella typhimurium* LT2 was transformed with the *FNT3* cloned into a pBAD vector (Invitrogen). Cells were adapted in MOPS medium supplemented with 10 mM glucose and thiamine at 1 µg ml⁻¹. A fraction of each culture (5 ml) was centrifuged and resuspended in fresh media supplemented with ampicillin, 0.2% arabinose and either 6 mM sodium sulphite or 6 mM sodium thiosulphate. Aliquots were incubated at 37 °C for 6 h and then centrifuged to pellet bacteria. Medium was gently poured into an equal volume of sulphide antioxidant buffer. Exported sulphide was measured using a combination silver sulphide electrode (Cole-Palmer). The concentration of sulphide in the medium was determined using standards of known sodium sulphide concentrations in the same antioxidant buffer.

Protein purification. *FNT3* was cloned into a modified pBAD vector (Invitrogen), creating a C-terminal TEV-FLAG-10×His fusion protein, and transformed into *E. coli* BL21 *plysS* (Sigma-Aldrich) cells^{11,34,35}. Transformed cells were grown at 37 °C to *D*_{600 nm} 1.0 and induced with 0.2% of arabinose for 4 h. Cells were collected and lysed by passing twice through an Emulsiflex cell disrupter (Avestin). Lysate was clarified by a low-speed spin at 12,000g and solubilized with 1% β-dodecyl-maltopyranoside (DDM, Anatrace). Solubilized lysate was incubated with Ni²⁺-NTA resin (Qiagen) and the bound protein was eluted with buffer containing 1.1% β-octyl-glucopyranoside (OG, Anatrace), 50 mM Tris 8.0, 200 mM NaCl, 300 mM imidazole and 10% glycerol. The protein sample was incubated with TEV protease overnight to remove the tags and then underwent size-exclusion chromatography in a Superdex 200 in 25 mM Tris, pH 6.8, 100 mM NaCl and 1.1% OG.

Mass spectrometry. The mass of purified HSC was determined using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in the laboratory of Dr T. Neubert following published protocols^{36–38}.

Protein-ion interaction assay. Ion-binding-induced thermostability of purified protein samples was measured using size-exclusion chromatography. Aliquots of 100 mg of Ni²⁺-NTA-purified protein were incubated for 10 min in a thermocycler at increasing temperatures. Samples were injected onto a Shodex KW804 analytical size-exclusion chromatography column (Thomson) on a high-performance liquid chromatograph (Shimadzu) equilibrated in 200 mM Na₂SO₄, 50 mM Tris, pH 8.0, 3 mM Na₃N and 0.05% DDM. The height of the monodisperse peak was used to quantify the amount of remaining sample. The nominal melting temperature, *T*_m, defined as the temperature at which 50% of the protein remained soluble, was calculated by fitting the data to a Boltzmann sigmoidal model in PRISM5 (GraphPad Software). For screening of various compounds, 20 mM of each sodium salt was added before incubation.

Proteoliposome reconstitution and *in vitro* transport assay. Transport of radiolabelled formate was measured in a concentrative uptake assay^{11,22}. *Escherichia coli* total lipids were aliquoted in chloroform and dried under nitrogen. Following resuspension in intraliposomal buffer (150 mM sodium formate and 50 mM Tris, pH 8.0), lipids were overlayed with nitrogen, freeze-thawed ten times,

vortexed and sonicated to homogeneity. Proteoliposomes were formed by the addition of 1% OG and Ni²⁺-NTA-purified HSC at a protein-to-lipid ratio of 1:10,000 (w/w). OG detergent was removed by incubation with 400 mg ml⁻¹ Bio-Beads SM2 (Bio-Rad) overnight. Proteoliposomes were then extruded through a 400-nm membrane filter and frozen at -20 °C until use. Concentrated re-uptake of formate was initiated by buffer-exchanging 200 ml of thawed proteoliposomes over 2 ml of G50 Sephadex swollen in extraliposomal buffer (150 mM glutamate, 50 mM Tris, pH 8.0). Carbon-14-labelled sodium formate (450 µM, American Radiolabelled Chemical) was added to the collected proteoliposomes, and the assays were terminated by centrifuging a 20-ml sample at 1,000g for 1 min at each time point through a G-50 Probequant Micro-column (GE Healthcare) to remove the external radiolabelled formate. The amount of transported radiolabelled formate was quantified by adding 1 ml of scintillation fluid to each sample and emitted photons were counted using a Wallac scintillation counter. For competition assays, increasing concentrations of sodium sulphite, sodium nitrite or sodium hydrosulphide (0, 0.15, 0.6, 3 and 15 mM), were added before the addition of radiolabelled formate.

Protein crystallization. High-pH, primitive orthorhombic crystals of HSC were grown using the vapour diffusion method, by mixing protein at a concentration of 11 mg ml⁻¹ in a 1:1 ratio with reservoir solution containing 25–27% (v/v) PEG 400, 100 mM Tris, pH 7.0–9.0, and 100 mM sodium sulphite, with or without 50 mM sodium hydrosulphide. Medium-pH, monoclinic crystals were grown in reservoir solution containing 25–27% (v/v) PEG 400, 100 mM Tris, pH 7.5, and 100 mM sodium nitrite. Low-pH, C-centred orthorhombic crystals were grown in buffer containing 100 mM sodium sulphite, 100 mM sodium nitrite, 8–10% (w/v) PEG 6000, 100 mM zinc acetate and 100 mM sodium acetate, pH 4.5.

Structure determination. X-ray diffraction data were collected at beamlines X25 and X29 of the National Synchrotron Light Source at Brookhaven National Laboratory and beamline 23ID of the Advanced Photon Source at Argonne National Laboratory. Images were processed and scaled using the HKL2000 suite³⁹. The structure was solved by molecular replacement using the *V. cholerae* FocA structure¹¹ (PDB code: 3KLY) as a search model. Models were built using PHENIX Autobuild⁴⁰ and manually adjusted using COOT⁴¹. Refinement was carried out using PHENIX Refine⁴⁰. The pore size of the channel was calculated using HOLE⁴², and structure figures were generated using PYMOL⁴³.

- Penel, S. *et al.* Databases of homologous gene families for comparative genomics. *BMC Bioinformatics* **10** (suppl. 6), S3 (2009).
- Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 (2004).
- Han, M. V. & Zmasek, C. M. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**, 356 (2009).
- Benson, D. A., Karsch-Mirzachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res.* **39**, D32–D37 (2011).
- Auer, M. *et al.* High-yield expression and functional analysis of *Escherichia coli* glycerol-3-phosphate transporter. *Biochemistry* **40**, 6628–6635 (2001).
- Wang, D. N. *et al.* Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. *Biochim. Biophys. Acta* **1610**, 23–36 (2003).
- Cadene, M. & Chait, B. A robust, detergent friendly method for mass spectrometry analysis of integral membrane proteins. *Anal. Chem.* **72**, 5655–5658 (2000).
- Li, X. D. *et al.* Monomeric state and ligand binding of recombinant GABA transporter from *Escherichia coli*. *FEBS Lett.* **494**, 165–169 (2001).
- Safferling, M. *et al.* The TetL tetracycline efflux protein from *Bacillus subtilis* is a dimer in the membrane and in detergent solution. *Biochemistry* **42**, 13969–13976 (2003).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276A**, 307–326 (1997).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Smart, O. S., Nedvědil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360, 376 (1996).
- DeLano, W. L. The PyMOL Molecular Graphics System (<http://www.pymol.org>) (2002).

Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å

Zachary Lee Johnson¹, Cheom-Gil Cheong¹ & Seok-Yong Lee¹

Nucleosides are required for DNA and RNA synthesis, and the nucleoside adenosine has a function in a variety of signalling processes^{1,2}. Transport of nucleosides across cell membranes provides the major source of nucleosides in many cell types and is also responsible for the termination of adenosine signalling. As a result of their hydrophilic nature, nucleosides require a specialized class of integral membrane proteins, known as nucleoside transporters (NTs), for specific transport across cell membranes. In addition to nucleosides, NTs are important determinants for the transport of nucleoside-derived drugs across cell membranes^{3–5}. A wide range of nucleoside-derived drugs, including anticancer drugs (such as Ara-C and gemcitabine) and antiviral drugs (such as zidovudine and ribavirin), have been shown to depend, at least in part, on NTs for transport across cell membranes^{4,6–13}. Concentrative nucleoside transporters, members of the solute carrier transporter superfamily SLC28, use an ion gradient in the active transport of both nucleosides and nucleoside-derived drugs against their chemical gradients. The structural basis for selective ion-coupled nucleoside transport by concentrative nucleoside transporters is unknown. Here we present the crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* in complex with uridine at 2.4 Å. Our functional data show that, like its human orthologues, the transporter uses a sodium-ion gradient for nucleoside transport. The structure reveals the overall architecture of this class of transporter, unravels the molecular determinants for nucleoside and sodium binding, and provides a framework for understanding the mechanism of nucleoside and nucleoside drug transport across cell membranes.

Humans have three isoforms of concentrative nucleoside transporter (hCNT), and the substrate specificities and tissue distributions of these isoforms are different^{14–16}. Knowledge of the mechanism of these transporters would help us not only to understand physiological processes associated with nucleosides but also to provide a framework for future drug design to improve nucleoside drug delivery. The major barrier to achieving a mechanistic understanding is the lack of atomic structures that reveal the origins of nucleoside specificity and the principles of function of CNTs. For structure determination and functional studies we chose a CNT homologue from *Vibrio cholerae* (vcCNT) because of its high sequence homology to hCNTs (39% identical to hCNT3; Supplementary Fig. 1) and optimal biochemical stability.

To test whether vcCNT can transport nucleosides, and if so what ion gradient vcCNT uses, we performed a radioactive nucleoside uptake assay using recombinant vcCNT-reconstituted liposomes. Studies have shown that the human CNTs use Na^+ ions and the *Escherichia coli* CNT uses H^+ for nucleoside transport^{17–19}. Transport activity was measured by monitoring the uptake of [5,6-³H]uridine in both the presence and the absence of a Na^+ gradient (Fig. 1a). We chose uridine because all of the CNTs that have been characterized so far transport uridine^{14,17–19}. When a Na^+ gradient was present, vesicles containing vcCNT accumulated radioactive uridine significantly more than the control empty vesicles. When a Na^+ gradient was not present (Na^+ was replaced with choline in the solution), uridine uptake was significantly decreased.

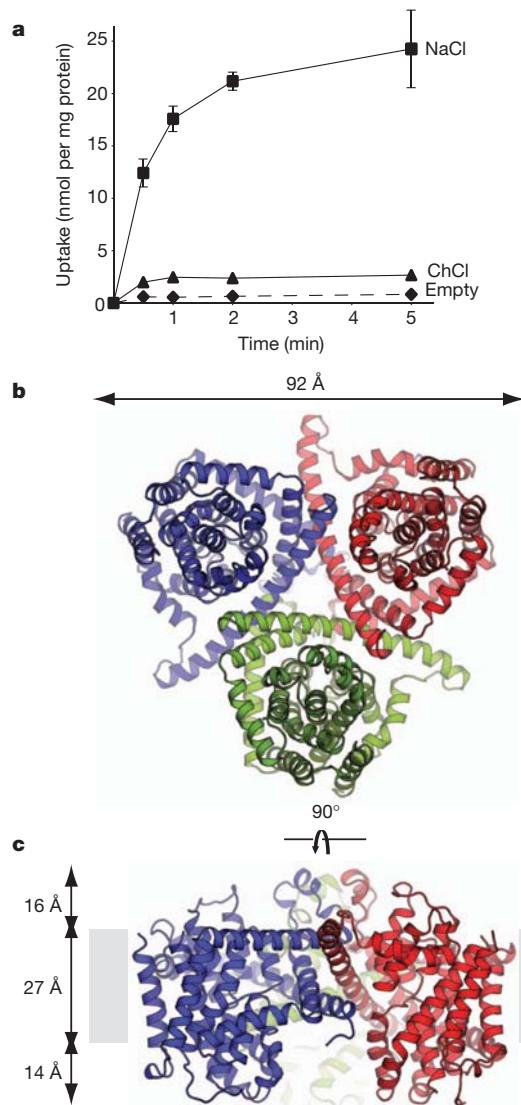


Figure 1 | vcCNT is a Na^+ -coupled nucleoside transporter with a trimeric architecture. **a**, Time course of the uptake of 2.4 μM [5,6-³H]uridine into vesicles containing vcCNT in the presence of a Na^+ gradient (squares, NaCl), in the absence of a Na^+ gradient (triangles, choline chloride), and the control empty vesicles in the presence of Na^+ (diamonds, empty). Results are means \pm s.d. ($n = 3$). **b**, Cartoon representation of the vcCNT trimer viewed from the cytoplasm. Individual protomers are coloured blue, red and green. **c**, Cartoon representation for the vcCNT trimer viewed parallel to the membrane. The putative membrane bilayer is indicated in **c** with grey boxes. The dimensions of the putative membrane bilayer, extracellular region, and intracellular region are shown.

¹Department of Biochemistry and Ion Channel Research Unit, Duke University Medical Center, 2 Genome Court, Durham, North Carolina 27710, USA.

Further characterization showed that uridine uptake was electrogenic and independent of pH gradients (Supplementary Fig. 2). These data clearly show that vcCNT, like human CNTs, uses a Na^+ gradient to transport nucleosides.

We crystallized and solved the structure of vcCNT at 2.4 Å (Supplementary Table 1). Experimental phases to 3.5 Å were obtained by single anomalous dispersion (SAD) from a platinum-soaked crystal (see Methods). The final model contains a single protomer in the asymmetric unit and is of good quality with a free *R* factor of 22.8% (Supplementary Table 1).

vcCNT crystallizes as a homotrimer that is shaped like an inverted triangular basin with its mouth facing the intracellular side and a knob-like structure facing the extracellular side (Fig. 1b, c). The three-fold axis coincides with the crystallographic six-fold axis and is perpendicular to the membrane. When viewed from the intracellular side, each side of the triangle formed by the trimer is about 92 Å, and when viewed parallel to the membrane the trimer is about 57 Å in height (Fig. 1). The membrane-embedded region lies roughly in the middle of the transporter, judging from the positions of three amphipathic helices on each protomer. On the basis of the predicted location of the membrane bilayer, the mouth of the basin penetrates into the membrane plane (Fig. 1c and Supplementary Fig. 3). The presence of several polar amino acids on the basin surface probably allows bulk aqueous solution to reach deep into the membrane bilayer and access this surface of the transporter.

The structure reveals that each protomer contains eight transmembrane helices (TM1–TM8), two re-entrant helix-turn-helix hairpins (HP1 and HP2) with opposite orientations in the membrane, and three interfacial helices (IH1–IH3) that run parallel to the membrane (Fig. 2a). The orientation of the structure relative to the

membrane is consistent with previous accessibility studies^{20–22} and conforms to the positive-inside rule (Supplementary Fig. 4). hCNTs are predicted to contain three more amino-terminal transmembrane helices than prokaryotic CNTs, suggesting an 11-TM topology for hCNTs²⁰.

Each protomer can be grouped into two subdomains on the basis of their locations (at the outer and inner regions) relative to the centre of the protomer. TM1, TM2, IH1, EH (a short stretch of extracellular helices), TM3 and TM6 are located at the outer part of the transporter, and they seem to be important for maintaining the overall architecture of the transporter. These helices in the outer region form a scaffold for the transporter (hereafter termed the scaffold domain) (Fig. 2). Trimerization contacts are mediated by part of the scaffold domain: IH1, EH, TM3 and TM6 (Fig. 2b). IH1 is an amphipathic helix 40 Å long that is most probably situated at the water/membrane interface, and TM6 is about 60 Å long and tilted almost 60° with respect to the membrane normal. EH protrudes into the extracellular solution and is about 16 Å in length. Because IH1 is involved in trimerization, is constrained to be at the membrane/water interface and interacts with many TMs (TM2–TM5), it is most probably important in building and maintaining the overall structure of vcCNT.

Surrounded by the scaffold domain, many conserved amino acids implicated in nucleoside transport are localized at the inner domain (hereafter termed the transport domain). The transport domain is composed of two structural groups that are related by an internal two-fold pseudo-symmetry. The first group is composed of IH2, HP1, TM4a/b and TM5 (pink background in Fig. 2a), and the second group is composed of IH3, HP2, TM7a/b and TM8 (cyan background in Fig. 2a). These two groups, separated by TM6 in sequence, can be superimposed with a root mean squared deviation of 3.3 Å, with the two-fold symmetry operator running through the middle of TM6 and

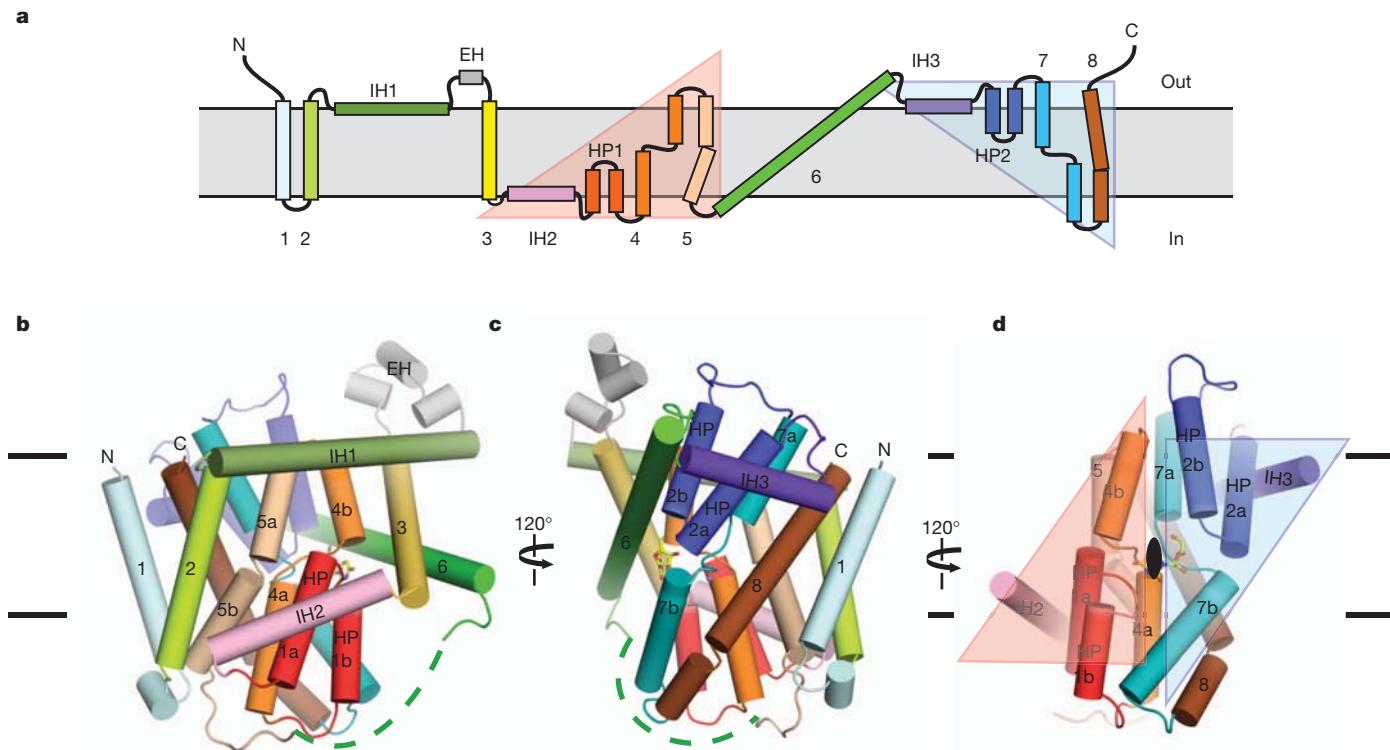


Figure 2 | Topology and fold of the vcCNT protomer. **a**, Schematic representation of vcCNT topology. The group of helices under the pink triangular background is related to the group of helices under the cyan triangular background by two-fold pseudo-symmetry, with the symmetry axis parallel to the membrane. **b**, Cartoon representation of the vcCNT protomer fold. Only helices comprising the scaffold domain and the group of helices under the pink triangular background are labelled. **c**, As in **b** but rotated through 120°. Only TM1, TM6 and helices comprising the group of helices under the cyan triangular background are labelled. **d**, As in **c** but rotated through 120°. The scaffold domain has been removed to show the pseudo-two-fold relationship between the two groups of helices, coloured as pink and cyan triangular backgrounds. The two-fold symmetry axis runs through the nucleoside at the centre of the transporter.

through 120°. Only TM1, TM6 and helices comprising the group of helices under the cyan triangular background are labelled. **d**, As in **c** but rotated through 120°. The scaffold domain has been removed to show the pseudo-two-fold relationship between the two groups of helices, coloured as pink and cyan triangular backgrounds. The two-fold symmetry axis runs through the nucleoside at the centre of the transporter.

parallel to the membrane (Fig. 2d). There is no significant amino-acid sequence homology between these two groups (roughly 10% sequence identity). This two-fold symmetry relationship positions the tips of HP1, HP2 and unwound regions of TM4 and TM7 at the centre of the transport domain, which is located slightly below the middle of the membrane plane (Fig. 2d). Sequence alignment of hCNTs and vcCNT reveals high sequence conservation around HP1, HP2 and the two unwound helices (TM4 and TM7), indicating the functional importance of this region (Supplementary Fig. 1).

To the best of our knowledge, the overall fold of vcCNT is novel, although local structural elements such as helical hairpins and unwound helices have been observed previously^{23,24}.

The crystal structure suggests that vcCNT adopts a trimeric configuration. To test whether the stoichiometry of vcCNT is trimeric and further validate the physiological relevance of our crystal structure, we performed structure-guided disulphide bridge crosslinking experiments (Supplementary Fig. 5). Our cysteine mutants readily form disulphide-crosslinked trimers in both detergent micelles and cell membranes under oxidizing conditions; therefore our crystal structure reflects a physiologically relevant oligomerization state. Given the sequence conservation of most of the amino acids involved in trimerization, we propose that the stoichiometry of both eukaryotic and prokaryotic CNT family members is trimeric (Supplementary Fig. 1).

The structure contains three deep clefts (one per subunit) at the intracellular side facing the centre of the trimer (Fig. 3a). A simulated-annealing OMIT map clearly shows that the electron density in the cleft is that of uridine (Fig. 3b). The bound uridine in the cleft faces the intracellular basin of the transporter; however, it is not free to be

released into the intracellular solution because TM6 and TM7b partly cover the binding site (Fig. 3a, b and Supplementary Fig. 6).

The nucleoside-binding site is located at the centre of the internal two-fold symmetry and lined by the tips of HP1 and HP2 and the unwound regions of TM4 and TM7 (Fig. 3b, c). Inspection of the interactions between uridine and the binding site shows that many polar or charged amino acids from HP1, HP2, TM4 and TM7 interact with the uracil base and ribose. HP1 and TM4b are responsible for interacting with the uracil base. The side chains of Gln 154, Thr 155 and Glu 156 from HP1 interact with the uracil base either directly (Gln 154) or indirectly through a water molecule (Thr 155 and Glu 156; Fig. 3c). Val 188 from TM4b interacts with the uracil base by means of van der Waals interactions. The involvement of these amino-acid residues in the interactions with the nucleoside is consistent with previous mutational studies^{25–27}: the residue corresponding to Gln 154, together with that corresponding to Val 188, is important for nucleoside specificity of hCNTs; Glu 156 is critical for Na^+ –nucleoside coupled transport.

HP2 and TM7 are responsible for the interactions with ribose. The side chains of Glu 332 (HP2), Asn 368 (TM7) and Ser 371 (TM7) interact with the ribose either directly (Glu 332, Asn 368 and Ser 371) or indirectly through a water molecule (Asn 368). Mutation of the residue corresponding to Glu 332 in hCNTs has been shown to have significant functional effects on both nucleoside binding and the rate of transport^{21,26}.

Because vcCNT is a Na^+ -coupled transporter, it must contain at least one Na^+ -binding site. Initial hints regarding the location of the Na^+ -binding site came from an $F_o - F_c$ map that shows a strong peak

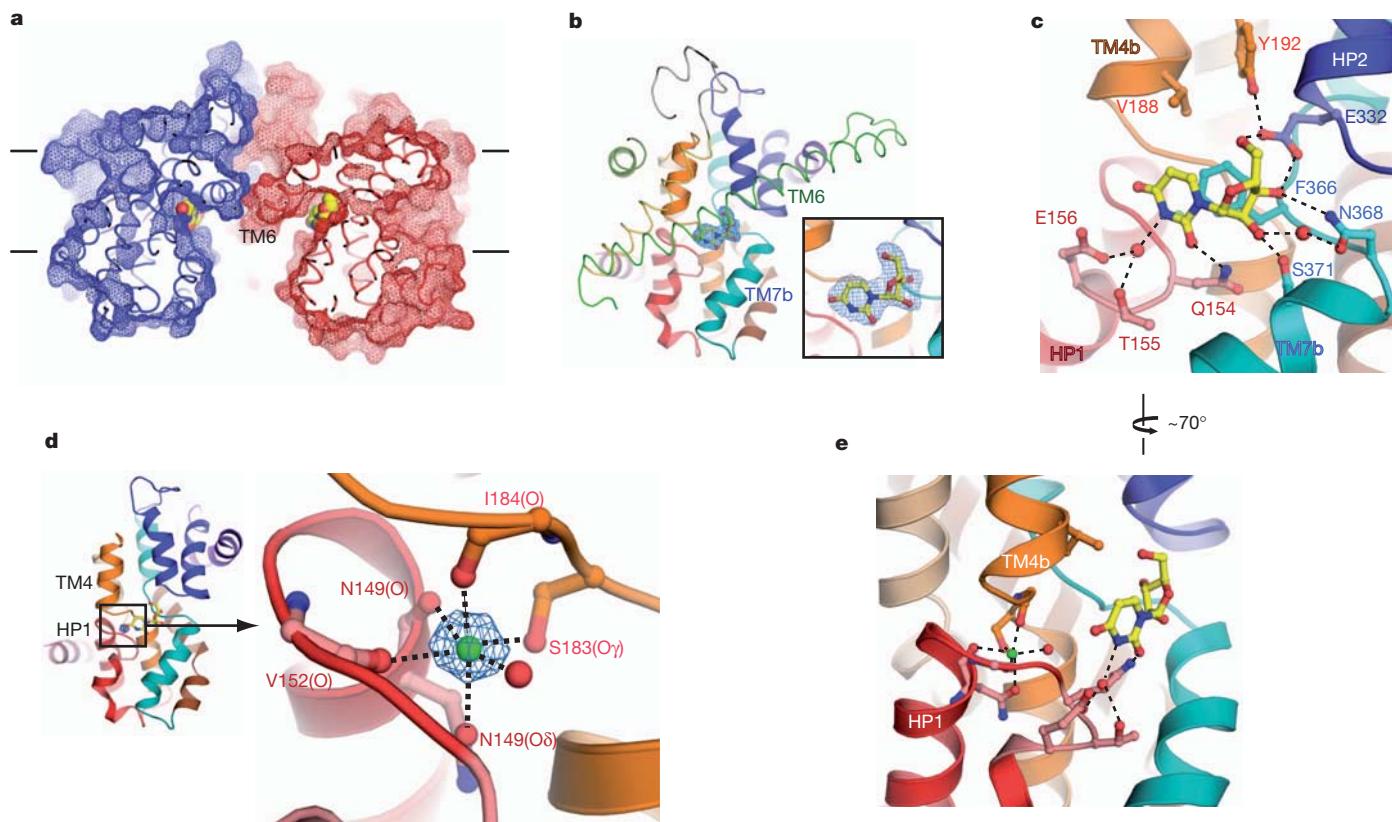


Figure 3 | Nucleoside-binding site and Na^+ -binding site. **a**, Cut-away surface representation of vcCNT viewed parallel to the membrane. The putative membrane bilayer is shown by horizontal lines. Uridine is shown as spheres. **b**, The vcCNT protomer viewed from the centre of the trimer. The scaffold domain is shown in ribbon representation; the transport domain is shown in cartoon representation. The blue mesh, covered by TM6 (green), is an $F_o - F_c$ simulated annealing OMIT map, contoured at 4σ , showing density for uridine.

The inset is a zoomed-in image of the simulated annealing electron density. **c**, The nucleoside-binding site, showing HP1 (red), TM4b (orange), HP2 (blue) and TM7b (cyan). Hydrogen bonds are shown as dashed lines. **d**, The Na^+ -binding site is located on HP1 (red) and the unwound region of TM4 (orange). The blue mesh is an $F_o - F_c$ simulated annealing OMIT map, contoured at 6σ , showing density for Na^+ . Coordination of the Na^+ ion is shown as dashed lines. **e**, The Na^+ -binding site is near the nucleoside-binding site.

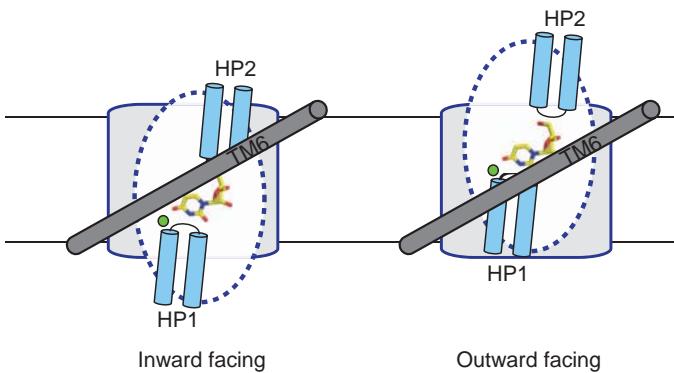


Figure 4 | Hypothetical mechanism of nucleoside transport. Only a single protomer is shown for simplicity, viewed from the centre of the trimer (same orientation as in Fig. 3b). On the left, the transporter adopts an inward-facing conformation. On the right, the transporter adopts an outward-facing conformation. Uridine is bound between the tips of HP1 and HP2, and TM6 (grey), serving as a hydrophobic barrier, partly covers uridine. The blue dotted lines demarcate the mobile part of the transport domain, and the grey region denotes the scaffold domain. Green circles represent sodium ions. A rigid-body motion of the nucleoside-binding site across TM6 can expose the binding site to the extracellular side.

(about 8σ) between the tip of HP1 and the unwound region of TM4. Further inspection reveals that this site is octahedrally coordinated by three backbone carbonyl groups, two side-chain hydroxyl groups and a water molecule (Fig. 3d) with the typical distances (about 2.4 Å) for Na^+ coordination by protein²⁸. Because it is difficult to distinguish Na^+ from H_2O on the basis of electron density alone, we performed the valence test that was used to identify the Na^+ sites in the crystal structure of LeuT^{28,29}. The result ($v = 1.02$) is consistent with Na^+ as the bound ion in this site. Therefore we suggest that vcCNT has at least one Na^+ -binding site at this position between HP1 and the unwound region of TM4.

The Na^+ -binding site is near the nucleoside-binding site, but the Na^+ does not interact directly with the nucleoside (Fig. 3e). What is the functional role of Na^+ binding at this site? Because key amino-acid residues for the interactions with the nucleoside base are localized on HP1 and TM4b, we propose that the role of Na^+ -binding is to bring HP1 close to TM4 to complete the formation of the nucleoside-binding site.

In the field of Na^+ -coupled secondary transporters, the prevailing model for ion-coupled transport is described by the alternating-access mechanism³⁰. If we assume that the alternating-access model is used, our crystal structure probably represents an inward-facing occluded conformation (Fig. 3a and Supplementary Fig. 6). On the basis of our crystal structure, previous accessibility studies of hCNTs, and studies of the Na^+ -coupled aspartate transporter Glt_{Ph}, which shares certain structural features with vcCNT (Supplementary Fig. 7), we propose a hypothetical model to explain Na^+ -coupled nucleoside transport by vcCNT (Fig. 4). A rigid-body movement of the transport domain across TM6 while the scaffold domain is held in place could permit the transition from an inward-facing to an outward-facing conformation. Our reasoning in proposing this mechanism is provided in Supplementary Information.

The crystal structure provides an important step towards understanding the mechanism of nucleoside transport by the CNT family. However, many important issues remain to be elucidated, including the structural basis for nucleoside and nucleoside-drug specificity and the conformational changes associated with each step along the transport cycle.

METHODS SUMMARY

vcCNT was expressed in *E. coli* C41 cells. Cells were lysed and protein was solubilized with dodecyl maltoside. Protein was purified by Co^{2+} -affinity chromatography followed by gel-filtration chromatography in the presence of decyl maltoside and uridine.

vcCNT was crystallized in the presence of 100 mM CaCl_2 , 37–42% PEG400 and 100 mM Tris-HCl pH 9.0 or 100 mM glycine pH 9.5. Platinum derivatives were prepared by soaking the crystals in $\text{K}_2\text{Pt}(\text{CNS})_6$. A partial poly-Ala model was built with phases to 5.0 Å from single isomorphous replacement with anomalous scattering (SIRAS). A complete model was built by combining phases from single anomalous dispersion using a new platinum derivative with molecular replacement phases from the partial model (MR-SAD) to 3.5 Å. Structure refinement was then performed against the 2.4-Å native data. The final model is of good quality with good Ramachandran statistics (98% favoured and 2% allowed). The model contains residues 2–416 (residues 230–240 missing), a uridine, 5 decylmaltosides (1 full and 4 partial), a Na^+ ion and 43 bound waters. Disordered residues Lys 226 and Glu 345 were modelled as Ala.

For the flux assay, protein was reconstituted into lipid vesicles and then vesicles were diluted into buffer containing either choline chloride or NaCl . Flux was initiated by the addition of 2.4 μM [5,6-³H]uridine and 1 μM valinomycin.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 August 2011; accepted 20 January 2012.

Published online 11 March 2012.

- King, A. E., Ackley, M. A., Cass, C. E., Young, J. D. & Baldwin, S. A. Nucleoside transporters: from scavengers to novel therapeutic targets. *Trends Pharmacol. Sci.* **27**, 416–425 (2006).
- Rose, J. B. & Cie, I. R. Physiology of nucleoside transporters: back to the future. *Physiology (Bethesda)* **23**, 41–48 (2008).
- Cano-Soldado, P. *et al.* Compensatory effects of the human nucleoside transporters on the response to nucleoside-derived drugs in breast cancer MCF7 cells. *Biochem. Pharmacol.* **75**, 639–648 (2008).
- Damaraju, V. L. *et al.* Nucleoside anticancer drugs: the role of nucleoside transporters in resistance to cancer chemotherapy. *Oncogene* **22**, 7524–7536 (2003).
- Jordheim, L. P. & Dumontet, C. Review of recent studies on resistance to cytotoxic deoxynucleoside analogues. *Biochim. Biophys. Acta* **1776**, 138–159 (2007).
- Errasti-Murugarren, E. & Pastor-Anglada, M. Drug transporter pharmacogenetics in nucleoside-based therapies. *Pharmacogenomics* **11**, 809–841 (2010).
- Mackey, J. R., Baldwin, S. A., Young, J. D. & Cass, C. E. Nucleoside transport and its significance for anticancer drug resistance. *Drug Resist. Updat.* **1**, 310–324 (1998).
- Mackey, J. R. *et al.* Immunohistochemical variation of human equilibrative nucleoside transporter 1 protein in primary breast cancers. *Clin. Cancer Res.* **8**, 110–116 (2002).
- Mackey, J. R. *et al.* Functional nucleoside transporters are required for gemcitabine influx and manifestation of toxicity in cancer cell lines. *Cancer Res.* **58**, 4349–4357 (1998).
- Nagai, K., Nagasawa, K. & Fujimoto, S. Uptake of the anthracycline pirarubicin into mouse M5076 ovarian sarcoma cells via a sodium-dependent nucleoside transport system. *Cancer Chemother. Pharmacol.* **55**, 222–230 (2005).
- Pastor-Anglada, M., Felipe, A. & Casado, F. J. Transport and mode of action of nucleoside derivatives used in chemical and antiviral therapies. *Trends Pharmacol. Sci.* **19**, 424–430 (1998).
- Pastor-Anglada, M. *et al.* Nucleoside transporters in chronic lymphocytic leukaemia. *Leukemia* **18**, 385–393 (2004).
- Zhang, J. *et al.* The role of nucleoside transporters in cancer chemotherapy with nucleoside drugs. *Cancer Metastasis Rev.* **26**, 85–110 (2007).
- Gray, J. H., Owen, R. P. & Giacomini, K. M. The concentrative nucleoside transporter family, SLC28. *Pflügers Arch.* **447**, 728–734 (2004).
- Molina-Arcas, M., Casado, F. J. & Pastor-Anglada, M. Nucleoside transporter proteins. *Curr. Vasc. Pharmacol.* **7**, 426–434 (2009).
- Ritzel, M. W. *et al.* Molecular identification and characterization of novel human and mouse concentrative Na^+ -nucleoside cotransporter proteins (hCNT3 and mCNT3) broadly selective for purine and pyrimidine nucleosides (system cib). *J. Biol. Chem.* **276**, 2914–2927 (2001).
- Loewen, S. K. *et al.* Functional characterization of a H^+ /nucleoside co-transporter (CaCNT) from *Candida albicans*, a fungal member of the concentrative nucleoside transporter (CNT) family of membrane proteins. *Yeast* **20**, 661–675 (2003).
- Loewen, S. K. *et al.* Transport of physiological nucleosides and anti-viral and anti-neoplastic nucleoside drugs by recombinant *Escherichia coli* nucleoside- H^+ cotransporter (NupC) produced in *Xenopus laevis* oocytes. *Mol. Membr. Biol.* **21**, 1–10 (2004).
- Xiao, G., Wang, J., Tangen, T. & Giacomini, K. M. A novel proton-dependent nucleoside transporter, CeCNT3, from *Caenorhabditis elegans*. *Mol. Pharmacol.* **59**, 339–348 (2001).
- Hamilton, S. R. *et al.* Subcellular distribution and membrane topology of the mammalian concentrative Na^+ -nucleoside cotransporter rCNT1. *J. Biol. Chem.* **276**, 27981–27988 (2001).
- Slugoski, M. D. *et al.* Substituted cysteine accessibility method analysis of human concentrative nucleoside transporter hCNT3 reveals a novel discontinuous region of functional importance within the CNT family motif (G/A)XKX3NEFVA(Y/M/F). *J. Biol. Chem.* **284**, 17281–17292 (2009).
- Zhang, J. *et al.* Cysteine-accessibility analysis of transmembrane domains 11–13 of human concentrative nucleoside transporter 3. *Biochem. J.* **394**, 389–398 (2006).

23. Cao, Y. *et al.* Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* **473**, 50–54 (2011).
24. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* **431**, 811–818 (2004).
25. Loewen, S. K. *et al.* Identification of amino acid residues responsible for the pyrimidine and purine nucleoside specificities of human concentrative Na^+ nucleoside cotransporters hCNT1 and hCNT2. *J. Biol. Chem.* **274**, 24475–24484 (1999).
26. Slugoski, M. D. *et al.* Conserved glutamate residues Glu-343 and Glu-519 provide mechanistic insights into cation/nucleoside cotransport by human concentrative nucleoside transporter hCNT3. *J. Biol. Chem.* **284**, 17266–17280 (2009).
27. Yao, S. Y. *et al.* Conserved glutamate residues are critically involved in Na^+ /nucleoside cotransport by human concentrative nucleoside transporter 1 (hCNT1). *J. Biol. Chem.* **282**, 30607–30617 (2007).
28. Nayal, M. & Di Cera, E. Valence screening of water in protein crystals reveals potential Na^+ binding sites. *J. Mol. Biol.* **256**, 228–234 (1996).
29. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^- -dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
30. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347–355 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Data for this study were collected at beamlines SER-CAT BM22/ID22 and NE-CAT ID 24-C at the Advanced Photon Source. We thank R. MacKinnon and J. Butterwick for critical reading; R. Lefkowitz and A. Shukla for providing access and technical support for the radioactive flux assay; S. Lockless for advice on experiments; and C. Pemble for help with remote data collection. This work was supported by start-up funds from the Duke University Medical Center, the McKnight Endowment Fund for Neuroscience, the Alfred P. Sloan Foundation, the Klingenstein Fund, the Mallinckrodt foundation, the Basil O'Connor Starter Scholar Research Award 5-FY10-473 from the March of Dimes Foundation, and the National Institutes of Health Director's New Innovator Award 1 DP2 OD008380-01 (all to S.-Y.L.).

Author Contributions Z.J. expressed, purified and crystallized vcCNT. Z.J. performed radioactive flux and crosslinking experiments. C.-G.C. participated in part of the vcCNT crystallization and generated mutants for crystallization and functional studies. Z.J. and S.-Y.L. collected and processed the data, solved the structure, and wrote the paper. S.-Y.L. designed the study. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structure are deposited in the Protein Data Bank under accession code 3TIJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.-Y.L. (sylee@biochem.duke.edu).

METHODS

Expression and purification. The gene encoding vcCNT was cloned from *Vibrio cholerae* genomic DNA into a modified pET26 vector that contains a pelB leader sequence and a PreScission Protease cleavable His₁₀-maltose-binding protein fusion. The original vector was a gift from R. Dutzler (University of Zurich) and was further modified. The vector containing vcCNT was expressed in *E. coli* C41 (DE3) cells. Cells were lysed with a homogenizer (Avestin) and protein was solubilized with 30 mM dodecyl maltoside (DDM) for 2 h at 4 °C. Solubilized lysates were spun down to remove the insoluble fraction, and supernatants were applied to Talon Co²⁺ affinity resin. After binding, protein was eluted with imidazole and digested overnight with PreScission Protease. The digestion mixture was concentrated and applied to a Superdex 200 size-exclusion column in the presence of 5 mM decyl maltoside (DM), 1 mM uridine, 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM dithiothreitol. Peak fractions corresponding to vcCNT were collected for further experiments.

Crystallization. Initial crystallization conditions were obtained at the high-throughput crystallization-screening laboratory at the Hauptman-Woodward Institute³¹. Protein was concentrated to about 10 mg ml⁻¹ and mixed 1:1 with crystallization solution containing 100 mM CaCl₂, 37–42% PEG400 and 100 mM buffer. Crystals grew over a wide pH range (5.6–9.5), but data for structure solution were collected on crystals grown at pH 9.0 (100 mM Tris-HCl) and pH 9.5 (100 mM glycine). Crystals were grown by using the microbatch-under-oil method. After 10–14 days, crystals were harvested, transferred to cryo solution containing 32.5% PEG400, and flash-frozen in liquid nitrogen. Platinum derivatives were prepared by soaking for 2–4 h in 2.5 mM K₂Pt(CNS)₆ and then transferring them to cryo solution and flash-freezing.

Structure determination. The data were collected on beamlines 22ID, 22BM and 24ID-C at the Advanced Photon Source. The data were processed with HKL2000 (ref. 32). Crystals of vcCNT diffract to 2.4 Å Bragg spacings and belong to the space group *P*6₃. Extensive screening of crystals was performed because roughly 70% of crystals are merohedrally twinned, with the twinning operator perpendicular to the crystallographic six-fold axis, leading to the apparent space group *P*6₃22. We found that SeMet-substituted crystals were almost always twinned with significant twinning fractions (20–45%), which necessitated heavy-atom-soaked crystals as the choice for a *de novo* phasing method for the vcCNT structure. Initial phases to 5.0 Å were obtained by single isomorphous replacement with anomalous scattering (SIRAS) from a platinum-soaked crystal. Platinum sites were found by SHELXD³³; phasing was calculated to 5.0 Å Bragg spacings by using SOLVE³⁴ with the figure of merit 0.34, and the density was modified by solvent flattening with RESOLVE³⁵. A partial model was built by manually placing idealized poly-Ala helices into the solvent-flattened electron density map by using COOT³⁶. After further extensive screening of platinum derivatives, we found a derivative that diffracted to 3.1 Å with significant anomalous signal but was non-isomorphous with any of our native data. Platinum sites were found from an anomalous difference Fourier map by using the partial model phases. Combined phases of single anomalous dispersion from the platinum derivative and molecular replacement

from the partial model (MR-SAD) were calculated at 4.0 Å with the figure of merit 0.35 and extended to 3.5 Å by solvent flattening with PHASER³⁷ and RESOLVE³⁵ with the use of the PHENIX interface³⁸. After iterative cycles of manual adjustment of poly-Ala helices and calculation of combined MR-SAD phases, the electron density map was of excellent quality and allowed us to place side chains into the partial model. We also collected native data to 2.4 Å at a long wavelength (1.6 Å) and identified sulphur sites by using an anomalous difference Fourier map, which helped guide the model building. After about 70% of manual model building was complete, molecular replacement was performed with the partial model against the 2.4 Å native data for further model building and refinement. Structure refinement was performed with PHENIX³⁸. An anomalous difference Fourier map with the native data to 2.4 Å collected at a long wavelength (1.6 Å) identified six sulphur sites and helped guide the model building. The final model is of good quality with *R*_{work}/*R*_{free} = 19.6/22.8% and good Ramachandran statistics (98% favoured and 2% allowed), and contains residues 2–416 (residues 230–240 missing), a uridine, 5 DMs (1 full and 4 partial), a Na⁺ ion and 43 bound waters.

Vesicle reconstitution and flux assay. Protein was reconstituted into lipid vesicles containing 10 mg ml⁻¹ of 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) and 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG) (3:1 ratio of POPE:POPG) at a mass ratio of 1:500 protein:lipid, as described previously³⁹. Vesicles were reconstituted in the presence of 200 mM KCl, 20 mM HEPES pH 7.4, 100 mM choline chloride. Reconstituted vesicles were then flash-frozen and thawed three times, then extruded through a 1.0-μm filter with the use of the Avanti Mini-Extruder.

For the flux assay, vesicles were diluted 1:20 into buffer containing 200 mM KCl, 20 mM HEPES pH 7.4, and either 100 mM choline chloride or 100 mM NaCl. Flux was initiated by the addition of 2.4 μM [5,6-³H]uridine and 1 μM valinomycin. All experiments were performed in triplicate at 30 °C. Vesicles were harvested on GF/B glass microfibre filters (Whatman) and counted by scintillation on the following day.

31. Koszelak-Rosenblum, M. *et al.* Determination and application of empirically derived detergent phase boundaries to effectively crystallize membrane proteins. *Protein Sci.* **18**, 1828–1839 (2009).
32. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
33. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
34. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861 (1999).
35. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
37. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
39. Lee, S. Y., Letts, J. A. & MacKinnon, R. Functional reconstitution of purified human H_v1 H⁺ channels. *J. Mol. Biol.* **387**, 1055–1060 (2009).

Collapse of polar ice sheets during the stage 11 interglacial

Maureen E. Raymo¹ & Jerry X. Mitrovica²

Contentious observations of Pleistocene shoreline features on the tectonically stable islands of Bermuda and the Bahamas have suggested that sea level about 400,000 years ago was more than 20 metres higher than it is today^{1–4}. Geochronologic and geomorphic evidence indicates that these features formed during interglacial marine isotope stage (MIS) 11, an unusually long interval of warmth during the ice age^{1–4}. Previous work has advanced two divergent hypotheses for these shoreline features: first, significant melting of the East Antarctic Ice Sheet, in addition to the collapse of the West Antarctic Ice Sheet and the Greenland Ice Sheet^{1–3}; or second, emplacement by a mega-tsunami during MIS 11 (ref. 4, 5). Here we show that the elevations of these features are corrected downwards by ~10 metres when we account for post-glacial crustal subsidence of these sites over the course of the anomalously long interglacial. On the basis of this correction, we estimate that eustatic sea level rose to ~6–13 m above the present-day value in the second half of MIS 11. This suggests that both the Greenland Ice Sheet and the West Antarctic Ice Sheet collapsed during the protracted warm period while changes in the volume of the East Antarctic Ice Sheet were relatively minor, thereby resolving the long-standing controversy over the stability of the East Antarctic Ice Sheet during MIS 11.

The stability of ice sheets in the face of continuing global warming is an issue of significant societal concern. Satellite gravity measurements indicate that the Greenland Ice Sheet (GIS) and the West Antarctic Ice Sheet (WAIS), the two ice sheets most susceptible to climate change, are experiencing a net mass loss^{6–9}, with evidence of an accelerating pace^{9–12}. In contrast, the current mass balance of the much larger East Antarctic Ice Sheet (EAIS) is uncertain, even in sign^{6,9}, though a recent study¹¹ has inferred EAIS mass loss localized to coastal regions. This uncertainty about the stability of the EAIS in a progressively warming world has been a key motivation for studies of the palaeoclimate record during past warm intervals.

One such study, a statistical analysis of widely distributed sea-level markers related to the last interglacial (MIS 5; about 120,000 years ago), concluded with 95% confidence that eustatic sea level (ESL; defined as the globally averaged sea-level change) was >6.6 m higher during MIS 5e than at the present day, and with 66% confidence that ESL was >8.0 m higher¹³. (This inference, higher than earlier estimates¹⁴, is supported by a recent analysis of MIS 5e sea-level records from Florida¹⁵.) Estimates of the ESL rise associated with collapse of polar ice sheets range from 3.4 m (ref. 16) to 7 m for the GIS, and from 3.2 m (ref. 17) to 5 m for the WAIS, where the upper bounds refer to the complete disappearance of the ice sheet. Thus, whereas the estimate of peak ESL during MIS 5e implies significant collapse of both the GIS and the WAIS, it also implies that the EAIS remained relatively stable.

It is within this context of assessing potential future instability of the EAIS that the sea-level highstand features found at ~20 m (here and elsewhere, height above present-day sea level is meant) in Bermuda and the Bahamas, and which formed during the MIS 11 interglacial (~424–395 kyr ago), have taken on great significance. MIS 11 spanned two precession cycles and was the longest interglacial of the past

500 kyr (refs 18, 19), including the current interglacial MIS 1 (Fig. 1) and MIS 5e (Supplementary Fig. 4). If the ESL during the MIS 11 interglacial peaked at a level 20 m higher than today^{1–3}, then at least 8 m of that rise must have come from melting of the EAIS. Geologic evidence for a ~20-m sea-level highstand in Bermuda and the Bahamas is convincing. In Bermuda, reasonably well-dated deposits with thalassinidean shrimp burrows, foraminifera, and gastropods characteristic of littoral and intertidal environments constrain relative sea level at 21.3 ± 1.0 m during MIS 11 (refs. 2, 3). On Eleuthera, in the Bahamas, a gently sloping erosion surface capped with fenestrae-rich intertidal beach deposits provides a maximum sea-level estimate of 20 ± 3 m, and the occurrence of pendant fibrous cements suggests a minimum sea level of 17 ± 2 m (we will henceforth quote a sea-level estimate of 18.5 ± 3.6 m for this site); multiple dating methods suggest that these deposits were formed during MIS 11 (ref. 1).

How do these observations compare to other MIS 11 sea-level indicators or proxies? In a recent survey of MIS 11 sea-level records worldwide (most of which are located in tectonically active regions), Bowen⁵ estimated peak MIS 11 sea level using a range of tectonic uplift

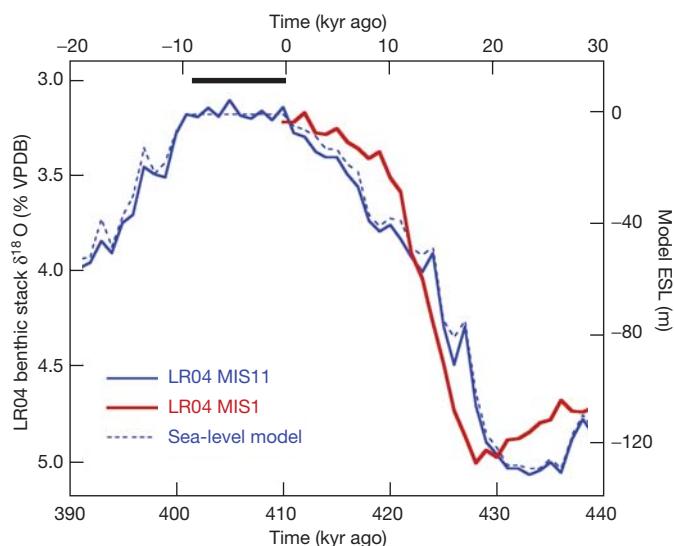


Figure 1 | Comparison of the duration of the MIS 11 and MIS 1 interglacials. Plot of the LR04 benthic oxygen isotope stack²⁸ (left-hand vertical axis) over a time window spanning the MIS 11 (blue; bottom time scale) and MIS 1 (red; top time scale) interglacials. The mean standard error on $\delta^{18}\text{O}$ in the LR04 stack is 0.06‰ with an age error of ± 4 kyr for the intervals considered here. The juxtaposition illustrates the significantly longer duration of maximum interglacial conditions during MIS 11 relative to MIS 1. ESL associated with the model ice history used to calculate GIA effects during MIS 11 is shown by dashed line (right-hand vertical axis). Note the hiatus in model ice volume changes from 410 to 401 kyr ago (black bar). An analogous comparison between the duration of MIS 11 with MIS 5e can be found in Supplementary Fig. 4.

¹Lamont-Doherty Earth Observatory, Columbia University, PO Box 1000, 61 Route 9W, Palisades, New York 10964, USA. ²Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA.

corrections based on different assumed ages and elevations of MIS 5e sea-level markers. If we adopt Kopp *et al.*'s¹³ recent inference of the age and peak magnitude of ESL during MIS 5e as a constraint on the latter, then Bowen's estimates from far-field sites imply that peak sea level during MIS 11 was 6–9 m above the present-day value, an estimate consistent with marine $\delta^{18}\text{O}$ isotopic anomalies²⁰. However, Bowen excluded a number of outliers, including the well-studied Bermuda and Bahamas sites, and following McMurtry *et al.*⁴, he attributed the anomalously high elevation of shoreline features at those sites to deposition by storms or a mega-tsunami.

The two explanations for the MIS 11 highstand features in Bermuda and the Bahamas—global sea-level rise associated with significant EAIS melting^{1–3}, or mega-tsunami deposition^{4,5}—ignore the potential signal from glacial isostatic adjustment (GIA). The GIA signal may be significant at the two sites, as they both are located on the peripheral bulge of the ancient Laurentian ice complex. As a consequence, they would be subject to local crustal subsidence and sea-level rise of amplitude $\sim 1\text{--}2 \text{ mm yr}^{-1}$ during any interglacial²¹.

To explore the potential contribution from GIA to MIS 11 highstand elevations, we compute global sea-level variations over the past 500 kyr using a gravitationally self-consistent theory valid for spherically symmetric, linear viscoelastic Earth models²² (see Supplementary Information). As an illustration of the physics of interglacial sea-level trends, we plot the predicted change in sea level across the warmest, most stable interval in MIS 11 (as implied by $\delta^{18}\text{O}$ records), a 9-kyr period spanning 410–401 kyr ago that occurred during the second half of MIS 11 (Fig. 2). The predictions adopt an Earth model with upper-mantle and lower-mantle viscosities of $5 \times 10^{20} \text{ Pa s}$ and $5 \times 10^{21} \text{ Pa s}$, respectively (model LM^{23,24}; see Supplementary Information).

In the near-field of the former MIS 12 ice sheets, the predicted sea-level change is dominated by radial crustal motions. For example, in regions once covered by ice sheets, post-glacial rebound of the crust produces a sea-level fall of amplitude up to $\sim 1\text{--}2 \text{ cm yr}^{-1}$, leading to a net fall over 9 kyr that can exceed 100 m (these amplitudes are well off the scale of Fig. 2). Surrounding these regions of uplift are peripheral bulges predicted to subside at rates of up to $0.2\text{--}0.3 \text{ cm yr}^{-1}$ during the interglacial. This subsidence accounts for the (red) zones of sea-level rise in Fig. 2, with maximum amplitude of $\sim 20 \text{ m}$, that encircle the Laurentian, Fennoscandian and Antarctic ice complexes. In contrast, a sea-level fall of 2–3 m is predicted across most of the far-field of the former ice sheets during MIS 11. This fall is due to a combination of two processes: (1) deglaciation-induced ocean loading effects, which act, near continental margins, to tilt the crust such that continents are deformed upward (that is, sea level falls) and offshore regions downward^{21,25}; and (2) the redistribution—or syphoning—of water from ocean basins towards regions of peripheral bulge and offshore subsidence²¹.

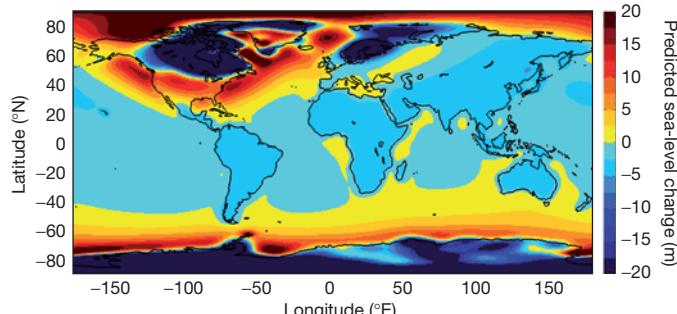


Figure 2 | Predicted sea-level change across the hiatus in ice volume changes (410–401 kyr ago) spanned by the model MIS 11 interglacial. The GIA calculation is based on the ice history discussed in Supplementary Information and the LM^{23,24} viscosity model. The colour scale saturates in regions within the near-field of the late Pleistocene ice sheets.

Following Fig. 2, we conclude that in the absence of interglacial ice volume changes and steric effects, sea-level highstands will date to the end of the interglacial at sites within the subsiding peripheral bulge of major ice centres, and to the beginning of the interglacial over most of the far-field (Fig. 3). (Of course, any sea-level signal associated with interglacial ice volume variations and with ocean temperature and salinity changes will be superimposed on these predicted GIA trends.) Consider Bermuda: the predicted relative sea-level history for this site (Fig. 3 and Supplementary Fig. 1, solid red line) exhibits a monotonic rise across MIS 11, reaching a level 11.9 m above the present-day value. The predicted highstand for the Bahamas (dashed red line, Fig. 3), lying on the outer flank of the same peripheral bulge, is 7.4 m. It is clear that the amplitude of highstands at Bermuda or the Bahamas, indeed at all sites within a peripheral bulge, will be a strong function of the duration of the interglacial highstand. In this regard, the protracted length of MIS 11, relative to the Holocene (Fig. 1) or MIS 5e (Supplementary Fig. 4), should be manifested by highstands of particularly large magnitude. From this it also follows that simple field comparisons of the relative elevations of Holocene, MIS 5e and MIS 11 sea-level markers on these islands would lead to an erroneous assessment of the difference in ESL between those times.

This result may be stated with more generality. The elevation of an ancient interglacial highstand at any site within a peripheral bulge will be governed by the difference in the state of isostatic disequilibrium at the end of the ancient interglacial relative to the disequilibrium at the present day in that same location. As the MIS 11 interglacial was significantly longer than the current interglacial (or MIS 5e), more subsidence had occurred by the time the final MIS 11 shoreline indicators were emplaced on those islands than during the current (or MIS 5e) interglacial. Further, as the current interglacial proceeds, sites like Bermuda and the Bahamas will be subject to a continuing sea-level rise as the present peripheral bulge continues to subside. Thus, the height of these ancient MIS 11 highstand features will progressively fall as time progresses (even in the absence of any future change in ice volume).

By contrast, the highstand amplitude at sites located in the far field will depend on the difference in isostatic disequilibrium at the beginning of the interglacial relative to the disequilibrium at the present day.

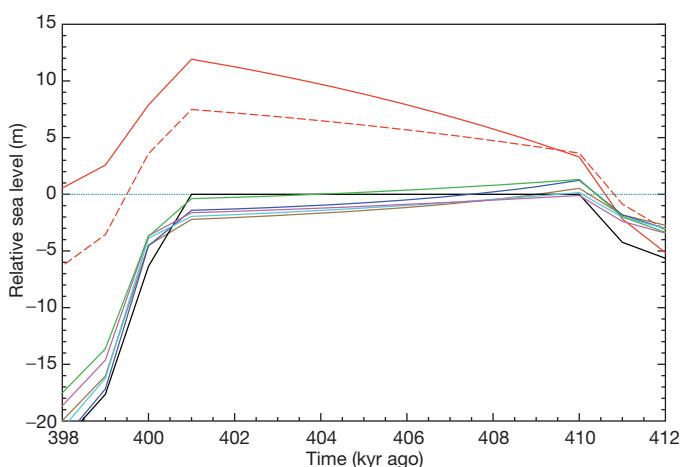


Figure 3 | Predicted relative sea-level changes across the model MIS 11 interglacial. The calculations are based on the ice history discussed in Supplementary Information and the LM viscosity model. The top two curves are predictions for Bermuda (solid red line) and the Bahamas (dashed red line) at locations of published data discussed in text. The lower coloured curves, predictions for a number of far-field sites, are included for comparison: the Gulf of Aden, Red Sea (brown), Curacao (green), Coorong, South Australia (magenta), Oahu, Hawaii (blue) and Cape Town, South Africa (cyan). The black line shows the ESL variation associated with the adopted ice history (as in the dashed blue line in Fig. 1 and the black line in Supplementary Fig. 1) and the thin horizontal dashed line references to present-day sea level.

Thus, perhaps counter-intuitively, the current elevation of these highstands does not depend on the duration of the ancient interglacial.

Global predictions of current MIS 11 shoreline elevations within each of the two zones discussed above are shown in Fig. 4a and b: Fig. 4a encompasses sites where the highstand occurred at the end of the modelled interglacial hiatus in ice mass change (for example, 401 kyr ago) whereas Fig. 4b encompasses sites where the highstand occurs at the beginning of the modelled hiatus (410 kyr ago). These results represent the predicted contribution to observed MIS 11 highstands from GIA alone; this contaminating signal should be removed from geological observations before they are used to infer the difference in ESL (or ice volume) between MIS 11 and the present.

As noted above, the GIA contribution to the MIS 11 shoreline elevations in Bermuda and the Bahamas is predicted to be 11.9 m and 7.4 m, respectively (Fig. 3). Correcting the published highstand estimates at these sites for these GIA predictions yields residual MIS 11 sea-level elevations of 9.4 ± 1 m for Bermuda and 11.1 ± 3.6 m in Eleuthera (Supplementary Table 1). Although additional field observations in the Bahamas could greatly decrease the multi-metre error associated with these sea-level estimates, we conclude that no significant contribution from melting of the EAIS is required to explain the geologic observations at the two sites.

In our analysis, we implicitly assume that melting of the WAIS and the GIS occurred towards the end of the MIS 11 interglacial, such that the contribution of the melt event to sea level at Bermuda and the Bahamas could be added to a coeval GIA-induced highstand. This timing for ice sheet collapse is strongly supported by at least two arguments. First, if the collapse took place at the beginning of the hiatus in MIS 11 melting (that is, at ~ 410 kyr ago), then the associated ESL rise would have had to be close to 20 m in order to fit the observed highstand elevations at Bermuda and the Bahamas. However, in this case, the predicted highstand elevation at far-field sites would have been even greater than 20 m (that is, an ~ 20 -m melt signal added to the GIA-induced highstand of a few metres; Fig. 3), a situation apparently ruled out by far-field observations⁵. Second, GIA calculations predict a submergence of up to several metres for far-field sites by 401 kyr ago (Fig. 3) and an average submergence of ~ 1 m at the specific far-field

sites considered by Bowen⁵. Therefore, correcting an inferred sea-level peak of 6–9 m for this GIA-induced contamination yields a predicted ESL ~ 1 m higher (that is, 7–10 m), in accord with the residual highstand elevations at Bermuda and the Bahamas cited above (Supplementary Table 1).

In the Supplementary Information we present results of sensitivity tests related to GIA predictions. First, we extend the duration of the MIS 11 hiatus in ice mass change from 9 to 14 kyr. In this case, the GIA-induced elevation of MIS 11 shorelines at sites within the Laurentian peripheral bulge are predicted to increase by ~ 1 –2 m when viscosity model LM is adopted, yielding residual MIS 11 highstand elevations of 7.0 ± 1 m for Bermuda and 9.9 ± 3.6 m in Eleuthera (Supplementary Table 1), and thus lowering the inferred peak ESL during MIS 11 by ~ 1 –2 m. Second, we considered an alternative viscosity profile, VM2 (ref. 26), characterized by a lower-mantle viscosity that is a factor of two smaller than model LM. In this case, the predicted GIA contribution to MIS 11 highstand elevations within the peripheral bulge is reduced by a factor of ~ 2 . This reduction yields residual (GIA-corrected) MIS 11 elevations of 15.0 ± 1 m and 14.3 ± 3.6 m at Bermuda and the Bahamas, respectively, values that appear to be at odds with lower far-field estimates⁵. Last, we performed a Monte Carlo parameter search in which we varied mantle viscosity, lithospheric thickness, and the duration of the modelled hiatus in MIS 11 ice volume changes, and tested for the consistency of the GIA-corrected highstand elevations. This exercise yields a preferred bound on the peak ESL during MIS 11 of 6–13 m.

In summary, observations of MIS 11 highstand features in Bermuda and the Bahamas must be corrected for the significant contaminating signal of GIA before these data are used to infer peak ESL during the MIS 11 interglacial. We conclude that ESL reached ~ 6 –13 m above the present-day value in the late stages of MIS 11. It is unlikely that ocean thermal expansion contributed more than ~ 1 m to this inference^{1,3,27}. Therefore, although this estimate of peak ESL indicates significant collapse of both the WAIS and GIS during the MIS 11 interglacial, it rules out any significant melting of the EAIS during this unusually prolonged period of ice age warmth.

Received 30 September 2011; accepted 23 January 2012.

Published online 14 March 2012.

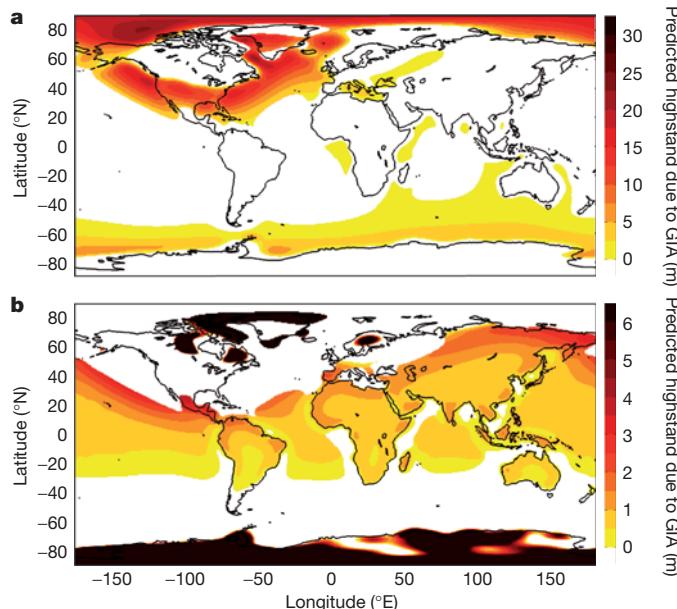


Figure 4 | Plot of the predicted GIA-induced present-day elevation of MIS 11 highstands that are currently greater than zero. Elevation is shown for all sites in which sea-level peaks either at the end of the hiatus in ice volume changes spanned by the model MIS 11 interglacial (a; that is, at 401 kyr ago; see also Supplementary Fig. 1) or at the beginning of that hiatus (b; that is, at 410 kyr ago).

1. Hearty, P. J., Kindler, P., Cheng, H. & Edwards, R. L. A +20 m middle Pleistocene sea level highstand (Bermuda and the Bahamas) due to partial collapse of Antarctic ice. *Geology* **27**, 375–378 (1999).
2. Olson, S. L. & Hearty, P. J. A sustained +21 m sea level highstand during MIS 11 (400 ka): direct fossil and sedimentary evidence from Bermuda. *Quat. Sci. Rev.* **28**, 271–285 (2009).
3. van Hengstum, P., Scott, D. B. & Javaux, E. Foraminifera in elevated Bermudian caves provide further evidence for +21 m eustatic sea level during Marine Isotope Stage 11. *Quat. Sci. Rev.* **28**, 1850–1860 (2009).
4. McMurtry, G. M. et al. Elevated marine deposits in Bermuda record a late Quaternary megatsunami. *Sedim. Geol.* **200**, 155–165 (2007).
5. Bowen, D. Q. Sea level ~ 400000 years ago (MIS 11): analogue for present and future sea-level? *Clim. Past* **6**, 19–29 (2010).
6. Velicogna, I. & Wahr, J. Measurements of time-variable gravity show mass loss in Antarctica. *Science* **311**, 1754–1756 (2006).
7. Luthcke, S. B. et al. Recent Greenland ice mass loss by drainage system from satellite gravity observations. *Science* **314**, 1286–1289 (2006).
8. Shepherd, A. & Wingham, D. Recent sea-level contributions of the Antarctic and Greenland ice sheets. *Science* **315**, 1529–1532 (2007).
9. Rignot, E. et al. Recent Antarctic ice mass loss from radar interferometry and regional climate modelling. *Nature Geosci.* **1**, 106–110 (2008).
10. Thomas, R. et al. Accelerated sea level rise from West Antarctica. *Science* **306**, 255–258 (2004).
11. Chen, J. L., Wilson, C. R., Blankenship, D. & Tapley, B. D. Accelerated Antarctic ice loss from satellite gravity measurements. *Nature Geosci.* **2**, 859–862 (2009).
12. Velicogna, I. Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE. *Geophys. Res. Lett.* **36**, L19503, <http://dx.doi.org/10.1029/2009GL040222> (2009).
13. Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C. & Oppenheimer, M. Probabilistic assessment of sea level during the last interglacial stage. *Nature* **462**, 863–867 (2009).
14. Jansen, E. et al. in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. et al.) 433–498 (Cambridge Univ. Press, 2007).

15. Muhs, D. R., Simmons, K. R., Schumann, R. R. & Halley, R. B. Sea-level history of the past two interglacial periods: new evidence from U-series dating of reef corals from south Florida. *Quat. Sci. Rev.* **30**, 570–590 (2011).
16. Otto-Bliesner, B. *et al.* Simulating Arctic climate warmth and ice field retreat in the last interglaciation. *Science* **311**, 1751–1753 (2006).
17. Bamber, J. L., Riva, R. E. M., Vermeersen, L. L. A., & LeBrocq, A. M. Reassessment of the potential sea level rise from a collapse of the West Antarctic Ice Sheet. *Science* **324**, 901–903 (2009).
18. Loutre, M. F. & Berger, A. Marine isotope stage 11 as an analogue for the present interglacial. *Glob. Planet. Change* **36**, 209–217 (2003).
19. Rohling, E. J. *et al.* Comparison between Holocene and Marine Isotope Stage-11 sea level histories. *Earth Planet. Sci. Lett.* **291**, 97–105 (2010).
20. McManus, J., Oppo, D., Cullen, J. & Healey, S. in *Earth's Climate and Orbital Eccentricity: The Marine Isotope Stage 11 Question* (eds Droxler, A. W., Poore, R. Z. & Burckle, L. H.) 69–85 (AGU Geophys. Monogr. Ser. 137, 2003).
21. Mitrovica, J. X. & Milne, G. A. On the origin of postglacial ocean syphoning. *Quat. Sci. Rev.* **21**, 2179–2190 (2002).
22. Kendall, R. A., Mitrovica, J. X. & Milne, G. A. On post-glacial sea level: II. Numerical formulation and comparative results on spherically symmetric models. *Geophys. J. Int.* **161**, 679–706 (2005).
23. Lambeck, K., Smith, C. & Johnston, P. Sea level change, glacial rebound and mantle viscosity for northern Europe. *Geophys. J. Int.* **134**, 102–144 (1998).
24. Mitrovica, J. X. & Forte, A. M. A new inference of mantle viscosity based upon a joint inversion of convection and glacial isostatic adjustment data. *Earth Planet. Sci. Lett.* **225**, 177–189 (2004).
25. Nakada, M. & Lambeck, K. Late Pleistocene and Holocene sea-level change in the Australian region and mantle rheology. *Geophys. J. Int.* **96**, 497–517 (1989).
26. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
27. McKay, N. P., Overpeck, J. T. & Otto-Bliesner, B. L. The role of ocean thermal expansion in Last Interglacial sea level rise. *Geophys. Res. Lett.* **38**, L14605, <http://dx.doi.org/10.1029/2011GL048280> (2011).
28. Lisicki, L. E. & Raymo, M. E. A. Pliocene–Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography* **20**, PA1003, <http://dx.doi.org/10.1029/2004PA001071> (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Hearty and D. Bowen for discussions of MIS 11 field data, and J. L. Davis for suggestions regarding data analysis. Support for this research was provided by NSF-OCE-0825293 and OCE-1202632 (M.E.R.), Harvard University (J.X.M.) and the Canadian Institute for Advanced Research (J.X.M.).

Author Contributions This study was planned, undertaken and written jointly.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.E.R. (raymo@ldeo.columbia.edu).

Suppression of the antiviral response by an influenza histone mimic

Ivan Marazzi¹, Jessica S. Y. Ho¹, Jaehoon Kim², Balaji Manicassamy^{3,4}, Scott Dewell⁵, Randy A. Albrecht^{3,4}, Chris W. Seibert³, Uwe Schaefer¹, Kate L. Jeffrey¹, Rab K. Prinjha⁶, Kevin Lee⁶, Adolfo García-Sastre^{3,4,7}, Robert G. Roeder² & Alexander Tarakhovsky¹

Viral infection is commonly associated with virus-driven hijacking of host proteins. Here we describe a novel mechanism by which influenza virus affects host cells through the interaction of influenza non-structural protein 1 (NS1) with the infected cell epigenome. We show that the NS1 protein of influenza A H3N2 subtype possesses a histone-like sequence (histone mimic) that is used by the virus to target the human PAF1 transcription elongation complex (hPAF1C). We demonstrate that binding of NS1 to hPAF1C depends on the NS1 histone mimic and results in suppression of hPAF1C-mediated transcriptional elongation. Furthermore, human PAF1 has a crucial role in the antiviral response. Loss of hPAF1C binding by NS1 attenuates influenza infection, whereas hPAF1C deficiency reduces antiviral gene expression and renders cells more susceptible to viruses. We propose that the histone mimic in NS1 enables the influenza virus to affect inducible gene expression selectively, thus contributing to suppression of the antiviral response.

Histones are essential regulators of genome function in eukaryotic cells^{1,2}. The amino-terminal domain of histones (the histone tail) provides a scaffold for the assembly of protein complexes controlling gene activity³. This role of a histone tail is achieved largely by post-translational histone modifications, catalysed by distinct modifying enzymes⁴. The combination of diversely modified histone tails establishes gene-region-specific patterns that, upon recognition by histone-binding proteins, contribute to the regulation of gene replication⁵, repair⁶ and transcription⁷.

The ability of histone tails to guide gene function indicates the possibility of targeted control of gene expression by artificial or naturally occurring molecules that can structurally and/or functionally mimic the histone tail. The former possibility has been demonstrated by studies that show the ability of synthetic compounds to interfere with inducible gene expression by abolishing the interaction between the acetylated histone H4 and the BET family of transcriptional regulators^{8,9}. Histone binding to transcriptional regulators could also be interrupted by exogenous cell-permeable histone peptides^{10,11}. This approach underscores the possibility of a competition between endogenous and exogenous histone tails for the common binding partners. Indeed, the histone H3 tail-like sequence (histone mimic) within histone methyltransferase G9a can compete, in a modification-dependent fashion, for binding to the histone-bound heterochromatin protein 1 (HP1)¹². Overall, the 3–5-amino-acid-long sequences that match various parts of the histone tail can be found in numerous eukaryotic and prokaryotic proteins; however, the role of these presumptive histone mimics in the regulation of gene activity remains unknown.

Influenza virus NS1 carries a histone H3-like sequence

Pathogens have a known ability to interfere with vital processes in the host cells by mimicking regulatory components of host protein networks¹³. In searching for naturally occurring histone mimics with a gene regulatory capability, we screened *in silico* for pathogen-derived proteins that have

a known or predicted capacity to accumulate within the nuclei of infected cells, and bear a histone-tail-like unstructured domain at the amino or carboxy termini. One of these proteins was found to be the influenza A virus NS1 protein that suppresses host response to influenza virus¹⁴.

Depending on the viral subtype, NS1 is a 219–237-amino-acid-long protein¹⁴. NS1 is not essential for formation of the viral particle but is critical for counteracting the antiviral cell response^{14,15}. In the absence of NS1, influenza is greatly attenuated¹⁵. Suppression of antiviral host response by NS1 relies partly on the ability of NS1 to interfere with cytosolic signalling processes that regulate the expression of type I interferon genes^{16–19}. Additionally, the NS1 protein can affect host gene expression by interfering with RNA splicing and messenger RNA export^{20–23}.

We found that NS1 of influenza A (H3N2) virus carries a sequence that resembles the histone H3 tail. Similar to the histone tail, the C terminus of NS1 comprises a non-structured and potentially highly interactive domain^{24,25} (Fig. 1a). Although this unstructured C-terminal domain is present in most of the human influenza A variants (Fig. 1a), only NS1 from H3N2 subtype (hereafter defined as NS1) possesses the ARSK sequence (amino acids 226–229) that is chemically analogous to the ¹ARTK⁴ sequence that comprises the lysine 4 site (H3K4) of histone H3 (Fig. 1a). The similarity between histone H3 and NS1 tails is further strengthened by the ability of the NS1 tail to serve as a substrate for histone-modifying enzymes *in vitro*. Incubation of the NS1 tail peptide with lysine methyltransferase Set1 complex or Set7/9 in the presence of S-adenosylmethionine resulted in incorporation of the methyl mark into the peptide (Fig. 1b, upper panel). Accordingly, substitution of lysine 229 to arginine (K229R) prevented methylation. These results, as well as the ability of histone acetyltransferase TIP60—which acetylates histone H3 lysine 4 in yeast²⁶—to acetylate the NS1 peptide at K229 (Fig. 1b, upper panel), support the histone mimicry by NS1. The ability of the H3K4-like sequence within the NS1 tail to serve as a substrate

¹Laboratory of Immune Cell Epigenetics and Signaling, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ²Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ³Department of Microbiology, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1124, New York, New York 10029, USA.

⁴Global Health and Infectious Pathogens Institute, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1124, New York, New York 10029, USA. ⁵Genomics Resource Center, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. ⁶Epinova DPU, Immuno-Inflammation Centre of Excellence for Drug Discovery, GlaxoSmithKline, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK. ⁷Department of Medicine, Division of Infectious Diseases, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1124, New York, New York 10029, USA.

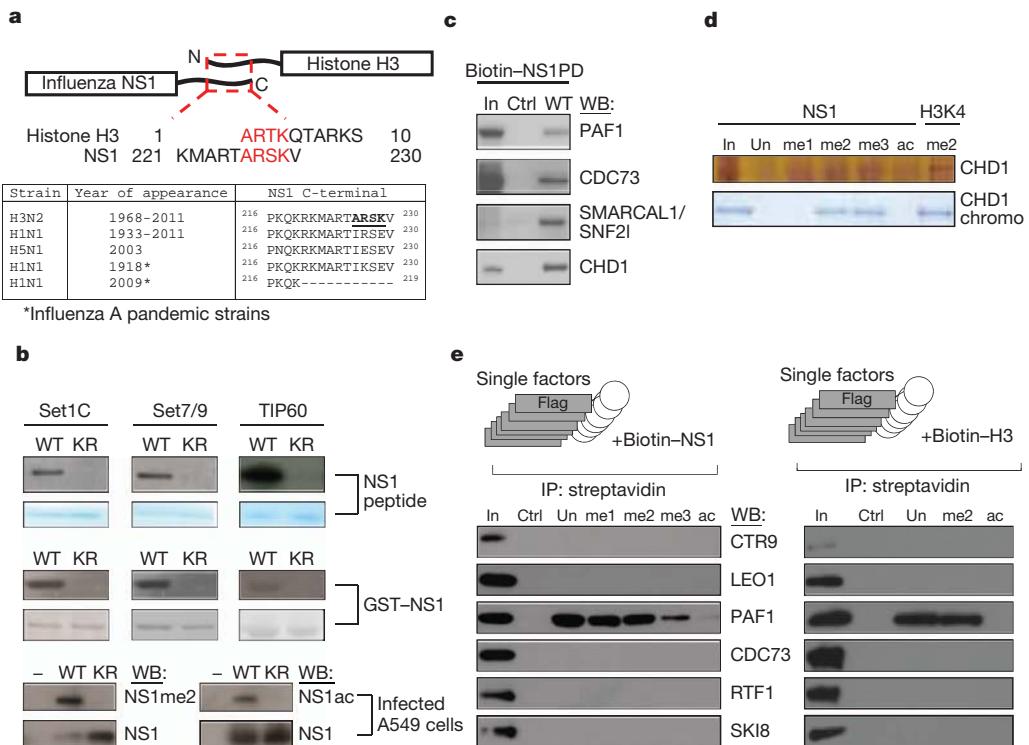


Figure 1 | Influenza NS1 contains a histone mimic. **a**, The homologous carboxy-terminal NS1 and the amino-terminal histone H3 sequences are shown (red letters). The table displays C-terminal NS1 sequences of the influenza A subtypes. **b**, Methylation or acetylation of the NS1 peptide (top panel), the GST-NS1 protein (middle panel) or of the viral NS1 in A549 infected cells (bottom panel) are shown. KR, NS1 substrates where K229 is replaced by arginine. **c**, Association of the NS1 histone mimic with the hPAF1C

for chosen histone-modifying enzymes has been further confirmed by using recombinant GST fusion NS1 protein (Fig. 1b, middle panel). Finally, the H3K4-like site within NS1 could be methylated or acetylated in virally infected cells. Using a custom-made antibody against dimethylated lysine 229 in NS1 (NS1me2; Supplementary Fig. 1), or acetyl-specific antibody²⁷, we found the presence of methylated and acetylated NS1 in infected human cells (Fig. 1b, lower panel). Point mutation of the NS1 gene that causes K229R substitution prevented modification (Fig. 1b, lower panel). It is likely that members of the H3K4 methyltransferase and acetyltransferase enzymes, that can modify NS1 *in vitro*, are responsible for NS1 modification *in vivo*. However, in view of the known redundancy between multiple H3K4 modifying complexes^{4,28–30}, identification of a sole NS1 modifying enzyme seems unlikely.

NS1 interacts with transcript-elongating proteins

Our experiments show the presence of a histone mimic within the NS1 tail. In histones, the unmodified or modified histone tail contributes to the assembly of chromatin protein complexes^{31–33}. The nature of the histone-tail-binding proteins is commonly revealed by the identification of nuclear proteins that bind to immobilized histone peptides during affinity purification³⁴. Using this approach (Supplementary Fig. 2), we searched for human nuclear proteins that bind to the NS1 tail peptide (amino acids 220–230). To control for the specificity of binding we used a scrambled peptide with an amino acid composition identical to NS1. Mass-spectrometry analysis of NS1-tail-bound proteins revealed the association of the NS1 tail with polypeptides that belong to the human PAF1 transcription elongation (hPAF1C) and CHD1 chromatin remodelling complexes (Supplementary Fig. 2). Binding of the NS1 tail to human PAF1, parafibromin (CDC73), SNF2I and CHD1 were validated by western blotting (Fig. 1c). In

subunits and CHD1 in nuclear extracts. In, input material. **d**, NS1 histone mimic binds to CHD1. Unmodified or methylated NS1 (K229) or methylated H3 (K4) peptides were incubated with the recombinant CHD1 or the CHD1 double-chromodomain. Binding to NS1 was revealed by silver or Coomassie staining (top and bottom panel, respectively). **e**, Binding of Flag-tagged hPAF1C subunits to NS1 or histone H3 peptides was assessed by western blotting (left and right panels, respectively). IP, immunoprecipitation.

support of the histone mimicry by the NS1 tail, we observed the histone-H3-like pattern of the NS1 tail binding to CHD1 *in vitro*. The chromodomain of human CHD1 binds specifically to di- or trimethylated lysine 4 of histone H3 (ref. 35). Similarly, the full-length CHD1 or the purified double-chromodomain binds selectively to di- or trimethylated NS1 peptide, but not to unmodified or acetylated NS1 tail (Fig. 1d).

To start assessing the role of the NS1 histone mimic in the regulation of chromatin complexes, we focused our attention on hPAF1C. We chose this on the basis of several factors. First, hPAF1C is a potent regulator of RNA elongation³⁶, a co-transcriptional process that contributes significantly to pathogen-induced gene expression^{11,37}. Second, PAF1C has been implicated in the regulation of stress-induced genes in yeast³⁸ that could be functionally equated to pathogen-induced genes in mammalian cells. Third, opposite to CHD1 that binds to the NS1 tail in a methyl-dependent fashion, purified hPAF1C binds to unmodified as well as modified NS1 tail (Supplementary Fig. 3a, left panel), thus making this interaction potentially more versatile. The interaction of hPAF1C with the NS1 tail is sequence-dependent. Neither NS1 tail of the H5N1 influenza virus, the scrambled NS1 peptide, nor the truncated NS1 tail that lacks the H3K4-like sequence were able to bind to hPAF1C (Supplementary Fig. 3b).

hPAF1C consists of six distinct subunits (PAF1, LEO1, CDC73 (also called parafibromin), SKI8 (also called WDR61), CTR9, RTF1) that have different roles in the interaction between hPAF1C and its binding partners, including RNA polymerase II (RNA Pol II)^{36,39,40}. To identify the primary NS1-tail-interacting hPAF1C subunit(s), we tested binding of the NS1 tail to purified recombinant individual hPAF1C subunits. We found that human PAF1 is the primary binder of the unmethylated or methylated NS1 tail, but not the control peptide that lacks the histone mimic (Fig. 1e, left panel). PAF1 does not bind to

acetylated NS1 tail, thus pointing to a possible *in vivo* regulation of NS1 binding to PAF1 by a methyl/acetyl-switch mechanism. Finally, NS1 interacts with PAF1 *in vivo*. Immunoprecipitation of NS1 or PAF1 from infected cell extract revealed the existence of an NS1-PAF1 complex (Supplementary Fig. 3c). Our findings show that direct binding of the NS1 histone mimic to PAF1 is responsible for the association of NS1 with hPAF1C.

Interaction of hPAF1C with the NS1 histone mimic raised the possibility of hPAF1C binding to the histone H3 tail. Although the interaction of hPAF1C with RNA Pol II and other transcriptional regulators is well established^{36,39,40}, the direct association between hPAF1C and histone H3 has not been documented. We found that unmodified and methylated histone H3 tail peptides, but not the scrambled control peptide, bind to hPAF1C (Supplementary Fig. 3a, right panel) or to purified PAF1 protein (Fig. 1e, right panel). Similar to the NS1 histone mimic, the acetylated histone H3 tail did not bind to either hPAF1C components.

The ability of homologous tails of NS1 and histone H3 to interact directly with hPAF1C indicated a possibility of virus-mediated hijacking of the transcription elongation machinery from the host genes. We found that NS1 accumulates in the nuclei of the infected human cells (Supplementary Fig. 4a), where it reaches approximately 5×10^5 molecules per nucleus within 12 h after infection (Supplementary Fig. 4b). The salt-extraction profile of NS1 from nuclei of infected cells showed association of NS1 with chromatin (Supplementary Fig. 4c). To assess the position of NS1 on chromatin during viral infection, we generated a recombinant virus (Flag-NS1) that expresses Flag-tagged NS1 protein (Supplementary Fig. 5a, b). Genome-wide analysis of the Flag-NS1 binding at 12 h after infection revealed the presence of NS1 on induced antiviral gene promoters (Supplementary Tables 1 and 2).

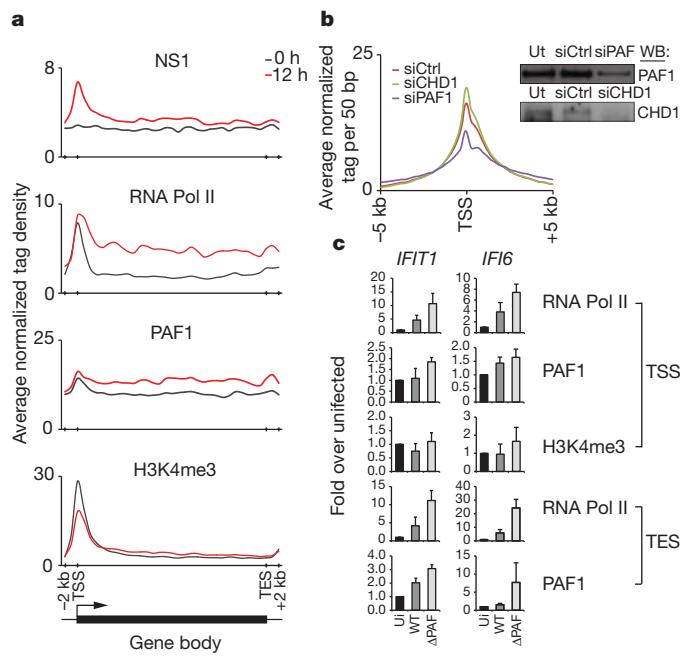


Figure 2 | Functional interaction between NS1 and PAF1 in infected cells. a, The ChIP-seq profiles show the distribution of indicated proteins at inducible genes before (black line) and after (red line) infection. The induced genes were revealed by RNA-seq and ChIP-seq analysis of infected A549 cells (Supplementary Tables 1 and 2). TSS and TES, the transcriptional start and end sites, respectively. b, The NS1 levels at gene promoters in PAF1- or CHD1-deficient cells (blue and green lines, respectively). The scrambled siRNA-treated cells (red line) were used as control. The insert shows knockdown of PAF1 or CHD1 in A549 cells. c, PAF1, RNA Pol II and H3K4me3 levels at the TSS and TES of the induced genes in uninfected (ui) cells, cells infected with the wild-type (WT) or PAF1-binding mutant virus (Δ PAF). Data are representative of three independent experiments; error bars show the s.e.m.

In infected cells, NS1 is enriched at promoters characterized by the presence of H3K4me3 and RNA Pol II (Fig. 2a). The pattern of NS1 distribution on its targets parallels the distribution of hPAF1C (Fig. 2a). Collectively, these data revealed a strategic position of NS1 at sites of active antiviral gene transcription.

NS1 histone mimic controls antiviral gene expression

The interaction between the NS1 tail and hPAF1C has two complementary functions that may support viral infection. The virus probably uses the NS1-hPAF1C interaction to target sites of active transcription. The short interfering RNA (siRNA)-mediated PAF1 deficiency resulted in 30% reduction of NS1 levels at the transcriptional start sites of the NS1-bound genes in infected cells (Fig. 2b). The specific role of PAF1 in recruitment of NS1 was supported by unaltered NS1 binding to its gene targets in cells treated with *CHD1*-specific siRNA (Fig. 2b). These findings revealed a previously unknown ability of the influenza virus to use host PAF1 to position viral NS1 protein at sites of active gene transcription.

Once targeted to transcriptionally active loci, as defined by RNA Pol II levels cross-referenced to the infected cell transcriptome (Supplementary Tables 1 and 2), NS1 interferes with PAF1 and RNA Pol II abundance at target genes. As exemplified by virus-induced *IFIT1* and *IFI6* genes, infection with the wild-type influenza virus resulted in decreased PAF1 and RNA Pol II levels at the transcriptional end sites (TES) and to a lesser extent at the transcriptional start sites (TSS) of the genes, as compared with PAF1-binding mutant virus (Δ PAF) (Fig. 2c). This pattern of PAF1, RNA Pol II and H3K4me3 distribution is indicative of impaired transcriptional elongation.

To determine the impact of the NS1 tail on the dynamics of antiviral gene transcription, we conducted a genome-wide nuclear 'run-on' analysis (global run-on sequencing, GRO-seq)⁴¹ of cells infected with the wild-type or mutant viruses that cannot bind to human PAF1 (Δ PAF). In the absence of infection, we observed accumulation of RNA species that map to the TSS of antiviral genes and low levels of RNAs that map to the gene body (Fig. 3a). Accumulation of TSS-associated short RNAs is a common feature of rapidly inducible genes that maintain a paused/poised state in the absence of an inducer⁴¹.

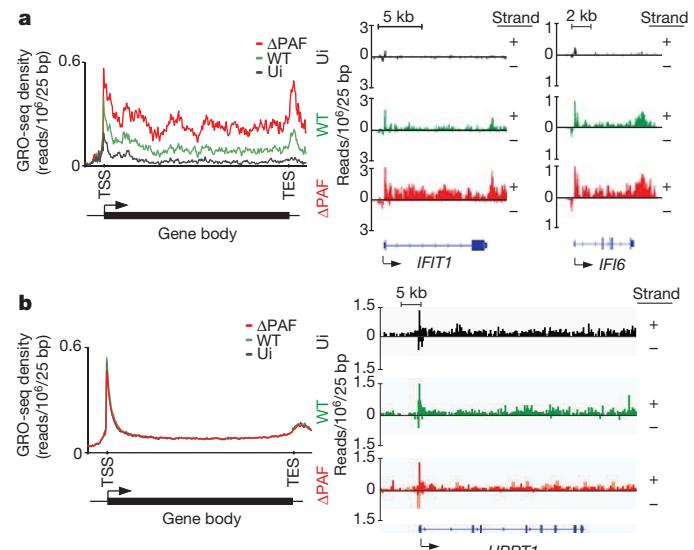


Figure 3 | NS1 suppresses antiviral gene transcription in infected cells. a, Left: the GRO-seq profile of inducible RNA transcripts in uninfected (ui) A549 cells (black line) or cells infected with wild-type or Δ PAF virus (green and red lines, respectively). Right: GRO-seq profile of *IFIT1* and *IFI6* genes in uninfected and infected cells. b, GRO-seq profile of A549-expressed genes that are not affected by virus infection (left panel) or of the *HPRT1* gene (right panel). Reads from either DNA strands are indicated as +/-. The y axes display reads per million mapped reads per 25 bp.

Infection with the Δ PAF virus resulted in a marked increase in levels of RNA transcripts that map to the TSS as well as to the downstream gene region. The high abundance of transcripts associated within the gene body is indicative of effective RNA elongation, which has an essential role in regulation of the pathogen-driven host response^{11,37}. Compared to cells infected with Δ PAF virus, the cells infected with wild-type virus displayed slightly reduced levels of TSS-associated RNAs but greatly diminished levels of transcripts associated with the gene body and 3' end region (Fig. 3a). This pattern is indicative of impaired transcriptional elongation and is consistent with defective elongation in other experimental settings^{42,43}. The differences between the impact of Δ PAF and wild-type viruses on the global dynamics of antiviral gene transcription are exemplified by differences in transcription of selected antiviral genes (Fig. 3a, right panel). Infection with either wild-type or Δ PAF virus has no effect on transcriptional dynamics of housekeeping genes and genes not affected by viral infection (Fig. 3b).

The ability of NS1 to control directly hPAF1C-dependent transcriptional elongation was further supported by the analysis of the impact of the NS1 tail on hPAF1C-driven RNA elongation *in vitro*. In this assay, a chromatinized DNA template, supplemented with selected transcription factors and histone-modifying enzymes, supports hPAF1C-mediated transcription in a fashion that faithfully reproduces transcriptional elongation *in vivo*³⁶ (Fig. 4a). We found that NS1 has a negative impact on hPAF1C function (Fig. 4b). The comparison between the amount of *in vitro* synthesized RNA in the absence or presence of hPAF1C shows the ability of hPAF1C to boost RNA synthesis (Fig. 4b). Addition of purified NS1 (Supplementary Fig. 5c) to the transcription reaction reduced the amount of synthesized RNA to background levels (Fig. 4b). Addition of purified NS1 that lacks the PAF1-binding sequence had no effect on RNA synthesis (Fig. 4b). Both wild-type NS1 and NS1 that lacks the PAF1 binding sequence have no impact on transcription of non-chromatinized template *in vitro* (Supplementary Fig. 6). This selective and direct impact of NS1 on virus-induced gene expression has been further supported by a lack of NS1 requirement for the association between the viral and the host RNA polymerases in infected cells^{44,45} (Supplementary Fig. 7).

Our findings provide unambiguous proof for the ability of the NS1 tail to interfere directly with hPAF1C-driven transcription of antiviral genes.

hPAF1C controls antiviral gene expression and infection

The significance of NS1–hPAF1C interaction during infection was evaluated by two independent experimental approaches. Deletion of the hPAF1C binding sequence in NS1 resulted in reduced viral titres of the Δ PAF virus as compared to those of the virus carrying full-length NS1 (Supplementary Fig. 8). This finding is in agreement with the previously reported positive role of the NS1 carboxy terminus in influenza A virulence⁴⁶.

We also revealed the central role of hPAF1C in transcriptional response induced by various viruses, as well as by the viral RNA analogue poly(I:C) and type I interferon. To study the impact of PAF1 on influenza-induced antiviral gene expression we used the

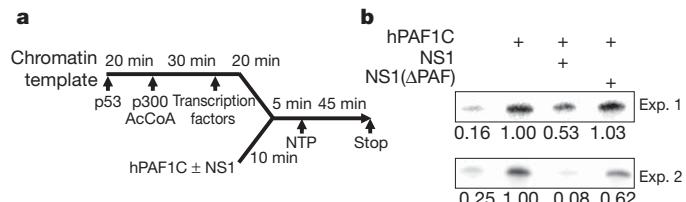


Figure 4 | NS1 inhibits transcriptional elongation *in vitro*. **a**, The full-length NS1 protein (NS1) or NS1 lacking the PAF1-binding sequence (NS1(Δ PAF)) (Supplementary Fig. 5c) was added to the RNA elongation reaction as indicated. **b**, The amount of the 390-nt RNA elongation product was quantified by ImageJ. The results of two independent experiments are shown.

NS1-deficient influenza virus (PR8/ Δ NS1) that elicits an extremely potent antiviral response and is therefore particularly useful to study host factors that regulate the antiviral response^{14,15}. Infection with PR8/ Δ NS1 of control siRNA-treated cells resulted in a strong upregulation of numerous antiviral genes (Fig. 5a). This response was reduced in PAF1-deficient cells (Fig. 5a, Supplementary Fig. 9, left panel, and Supplementary Table 3). The negative effect of PAF1 deficiency on antiviral response is not a consequence of impaired virus life-cycle caused by a reduction in viral gene expression (Supplementary Fig. 10).

PAF1 deficiency also attenuated the transcriptional response to wild-type influenza A H1N1 (Supplementary Fig. 11, left panel). The general positive role of PAF1 in antiviral gene expression was underscored by reduced expression of antiviral genes in PAF1-deficient cells infected with vesicular stomatitis virus (VSV) as well as cells treated with poly(I:C) (Supplementary Fig. 11 and Supplementary Tables 4–6). PAF1 deficiency also downregulates the transcriptional response to type I interferon that governs the antiviral immunity (Fig. 5b, Supplementary Fig. 9, right panel and Supplementary Table 7). The impaired antiviral response by PAF1-deficient cells contributes to a nearly tenfold increase of NS1-deficient PR8/ Δ NS1 virus replication,

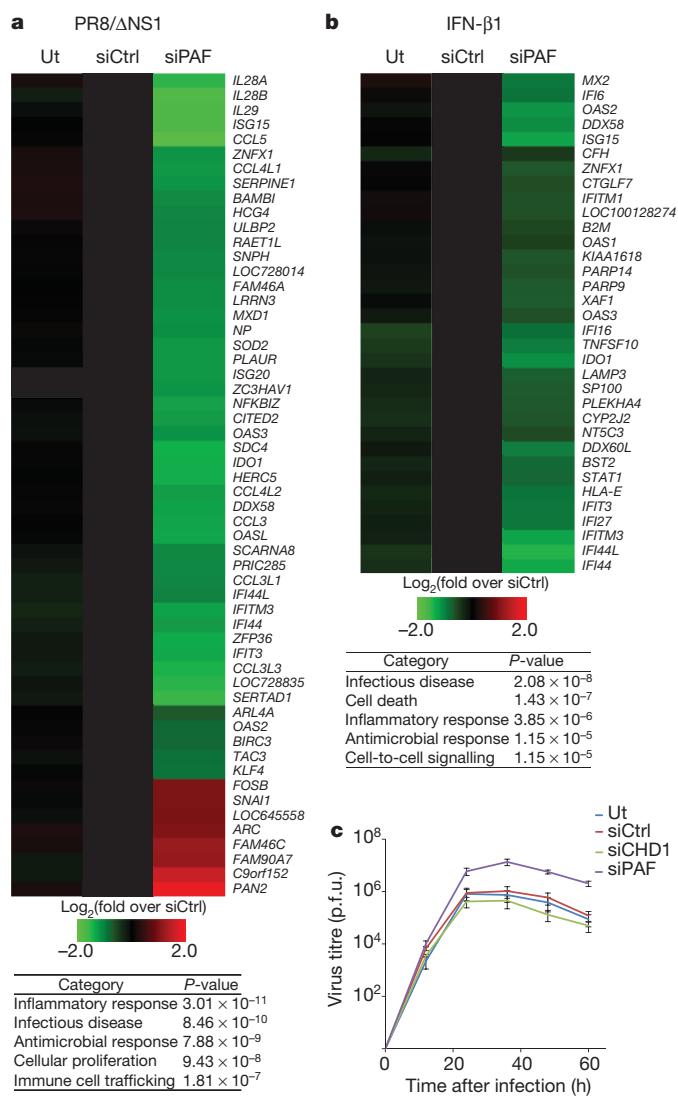


Figure 5 | PAF1 controls antiviral response. **a, b**, The expression levels of mRNAs in influenza infected (a) or IFN- β 1-treated (b) control (siCtrl) or PAF1-deficient (siPAF) A549 cells. The tables show the siPAF-affected gene categories. Ut, untreated with siRNA. **c**, Dynamics of virus replication in control or PAF1-deficient A549 cells. p.f.u. plaque-forming units. Data are representative of three independent experiments. Error bars show the s.e.m.

as compared to infected control cells (Fig. 5c). The impact of PAF1 deficiency on antiviral gene expression is selective. PAF1 deficiency does not alter the expression of housekeeping genes (Supplementary Fig. 12 and Supplementary Table 8).

Concluding remarks

We have shown that H3N2 influenza A virus interferes with host gene expression by exploiting the very basic principles of the epigenetic control of gene regulation. By mimicking the histone H3K4 sequence, which has a key role in positive regulation of gene transcription, the influenza virus gains access to histone-interacting transcriptional regulators that govern inducible antiviral gene expression. The histone-like unstructured NS1 tails are present in the majority of human influenza A isolates, but their primary sequence differs between viruses. The presence of potentially reactive amino acid residues within the unstructured NS1 tails and the nuclear localization of NS1 predict a multitude of nuclear proteins interacting with influenza NS1 in a sequence- and/or modification-specific fashion. In extension of the histone code that has provided a conceptual framework for understanding chromatin-based gene regulation^{47–49}, we propose the existence of the ‘NS1 code’ that guides influenza virus interference with host chromatin. The interference with chromatin function may complement other immunosuppressive functions of NS1 and provide the influenza virus with an opportunity to affect host gene expression in a highly selective fashion by recognizing and using epigenetic patterns of the host cells. This mechanism may contribute to the influenza-variant-specific pathogenesis and disease progression that differ between viruses carrying the histone-like NS1 tail and those viruses, such as human H1N1 (including pandemic 2009 virus), that lack the NS1 tail⁵⁰. It is possible that by targeting the actively transcribed antiviral genes, the NS1 tail helps influenza to balance its virulence against the preservation of the infected cell function. Such a mechanism could help the virus to maintain its long-term presence within the human population. Our findings could also be applied to the development of novel anti-influenza therapies that will interfere in a highly selective fashion with the NS1 tail binding to its chromatin partners. Finally, identification of hPAF1C as a key mediator of antiviral and pro-inflammatory gene response may pave the way for the development of anti-inflammatory therapies that—similar to the recently described ‘epigenetic’ therapies^{8,9}—will interfere selectively with hPAF1C binding to its partners.

METHODS SUMMARY

The methylation and acetylation *in vitro* has been performed as described previously^{12,26}. The analysis of NS1 methylation or acetylation *in vivo* has been done by using custom-made methyl-specific or commercial acetyl-specific antibodies²⁷. Isolation of the NS1 binding proteins, their analysis by mass-spectrometry and western blotting were performed according to a protocol used to identify histone-bound proteins³⁴. The recombinant hPAF1C and its individual components were prepared as described before³⁶. Interaction between hPAF1C and its individual components with peptides were performed as reported^{32,33}. The transcriptional assays have been performed as described³⁶. Genome-wide distribution of NS1 and other chromatin-bound proteins were accessed by ChIP-sequencing as previously described⁹, and analysed using standard bioinformatic approaches. Recombinant influenza viruses were generated using standard reverse genetics approaches^{15,17}. Infection of cells with viruses was performed according to the standard methods^{17,46} and viral titres were quantified by plaque assay. Gene expression analysis has been done by microarray using Illumina Human HT-12 v4 Expression BeadChips or by RNA-sequencing and verified by quantitative real-time PCR⁹. GRO-sequencing was performed as described⁴¹. All methodological details are additionally outlined in Supplementary Information.

Received 7 September 2011; accepted 23 January 2012.

Published online 14 March 2012.

1. Kornberg, R. D. & Thomas, J. O. Chromatin structure—oligomers of histones. *Science* **184**, 865–868 (1974).
2. Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Ann. Rev. Genet.* **43**, 559–599 (2009).

3. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
4. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
5. Kelly, A. E. *et al.* Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. *Science* **330**, 235–239 (2010).
6. Fernandez-Capetillo, O. *et al.* DNA damage-induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biol.* **4**, 993–997 (2002).
7. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
8. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
9. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119–1123 (2010).
10. Nishiyama, A. *et al.* Intracellular delivery of acetyl-histone peptides inhibits native bromodomain-chromatin interactions and impairs mitotic progression. *Febs Lett.* **582**, 1501–1507 (2008).
11. Hargreaves, D. C., Horng, T. & Medzhitov, R. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138**, 129–145 (2009).
12. Sampath, S. C. *et al.* Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol. Cell* **27**, 596–608 (2007).
13. Elde, N. C. & Malik, H. S. The evolutionary conundrum of pathogen mimicry. *Nature Rev. Microbiol.* **7**, 787–797 (2009).
14. Hale, B. G., Randall, R. E., Ortin, J. & Jackson, D. The multifunctional NS1 protein of influenza A viruses. *J. Gen. Virol.* **89**, 2359–2376 (2008).
15. Garcia-Sastre, A. *et al.* Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. *Virology* **252**, 324–330 (1998).
16. Lu, Y., Wambach, M., Katze, M. G. & Krug, R. M. Binding of the influenza-virus NS1 protein to double-stranded-RNA inhibits the activation of the protein-kinase that phosphorylates the Elf-2 translation initiation-factor. *Virology* **214**, 222–228 (1995).
17. Gack, M. U. *et al.* Influenza A Virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe* **5**, 439–449 (2009).
18. Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997–1001 (2006).
19. Hale, B. G., Jackson, D., Chen, Y. H., Lamb, R. A. & Randall, R. E. Influenza A virus NS1 protein binds p85b and activates phosphatidylinositol-3-kinase signaling. *Proc. Natl. Acad. Sci. USA* **103**, 14194–14199 (2006).
20. Krug, R. M., Yuan, W. M., Noah, D. L. & Latham, A. G. Intracellular warfare between human influenza viruses and human cells: the roles of the viral NS1 protein. *Virology* **309**, 181–189 (2003).
21. Nemeroff, M. E., Barabino, S. M. L., Li, Y. Z., Keller, W. & Krug, R. M. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. *Mol. Cell* **1**, 991–1000 (1998).
22. Das, K. *et al.* Structural basis for suppression of a host antiviral response by influenza A virus. *Proc. Natl. Acad. Sci. USA* **105**, 13093–13098 (2008).
23. Satterly, N. *et al.* Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proc. Natl. Acad. Sci. USA* **104**, 1853–1858 (2007).
24. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
25. Hale, B. G., Barclay, W. S., Randall, R. E. & Russell, R. J. Structure of an avian influenza A virus NS1 protein effector domain. *Virology* **378**, 1–5 (2008).
26. Xhermalce, B. & Kouzarides, T. A chromodomain switch mediated by histone H3 Lys 4 acetylation regulates heterochromatin assembly. *Genes Dev.* **24**, 647–652 (2010).
27. Becker, P. B. *et al.* Site-specific acetylation of ISWI by GCN5. *BMC Mol. Biol.* **8** (2007).
28. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15–30 (2007).
29. Wang, P. F. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29**, 6074–6085 (2009).
30. Guillemette, B. *et al.* H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. *PLoS Genet.* **7** (2011).
31. Lachner, M., O'Carroll, N., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
32. Shi, X. B. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).
33. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**, 718–722 (2007).
34. Wysocka, J. Identifying novel proteins recognizing histone modifications using peptide pull-down assay. *Methods* **40**, 339–343 (2006).
35. Sims, R. J. *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280**, 41789–41792 (2005).
36. Kim, J., Guermah, M. & Roeder, R. G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
37. Ramirez-Carrozzi, V. R. *et al.* A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138**, 114–128 (2009).

38. Kim, K. Y. & Levin, D. E. Mpkl MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell* **144**, 745–756 (2011).
39. Chen, Y. X. *et al.* DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev.* **23**, 2765–2777 (2009).
40. Jaehning, J. A. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim. Biophys. Acta* **1799**, 379–388 (2010).
41. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
42. Min, I. M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Gene Dev.* **25**, 742–754 (2011).
43. Mapendano, C. K., Lykke-Andersen, S., Kjems, J., Bertrand, E. & Jensen, T. H. Crosstalk between mRNA 3' end processing and transcription initiation. *Mol. Cell* **40**, 410–422 (2010).
44. Loucaides, E. M. *et al.* Nuclear dynamics of influenza A virus ribonucleoproteins revealed by live-cell imaging studies. *Virology* **394**, 154–163 (2009).
45. Engelhardt, O. G., Smith, M. & Fodor, E. Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *J. Virol.* **79**, 5812–5818 (2005).
46. Jackson, D., Hossain, M. J., Hickman, D., Perez, D. R. & Lamb, R. A. A new influenza virus virulence determinant: The NS1 protein four C-terminal residues modulate pathogenicity. *Proc. Natl Acad. Sci. USA* **105**, 4381–4386 (2008).
47. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
48. Turner, B. M. Histone acetylation and an epigenetic code. *Bioessays* **22**, 836–845 (2000).
49. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
50. Yang, Y. *et al.* The transmissibility and control of pandemic influenza A (H1N1) virus. *Science* **326**, 729–733 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. deGross and A. Rudensky for the mass spectroscopy analysis of the NS1 binding proteins. A. Rojas Soto, D. Reinberg, M. Dobenecker and T. Zhanyun provided us with recombinant CHD1 (A.R.S., D.R.), recombinant Set7/9 (M.D.) and Set1C (T.Z.). F. Casadio, P. Lewis, O. Binda, O. Gozani, N. Levenkova, A. Mele, R. Darnell, L. Core, J. Lis and P. Palese gave us valuable technical advice and help with data analysis. We acknowledge the Rockefeller University Genomics Resource Center for technical support. We thank R. Cadagan, A. Santana, W. Huang, R. Chandramouli and H. Zebronsky for technical assistance, R. Rizzo for help with manuscript preparation and C. Nathan for discussion. L.M.K. for artwork. B.M. is supported by NIH/NIAID K99 Pathway to Independence award (1K99AI095320-01). A.G.-S. is partially supported by NIAID grants R01AI046954, U19AI083025 and by CRIP (Center for Research in Influenza Pathogenesis), an NIAID funded Center of Excellence for Influenza Research and Surveillance, HHSN266200700010C. R.G.R. is supported by NIH grant CA129325. J.K. is supported by Charles H. Revson Foundation. I.M. is supported by American Italian Cancer Foundation. J.H. is supported by the Agency for Science, Technology and Research (A*STAR), Singapore. A.T. is supported by the NIH grant R01AI068058 and by Starr Cancer Consortium.

Author Contributions I.M. contributed to design, execution, analysis of the experiments and manuscript preparation. J.S.Y.H. studied the role of PAF1 in viral infection and assisted in manuscript preparation. J.K. and R.R. studied the impact of NS1 on hPAF1C and transcriptional elongation. B.M., R.A.A. engineered the recombinant influenza viruses and studied viral infectivity. U.S. was involved in gene expression studies. S.D. performed bioinformatic analysis. C.W.S. generated antibody against viral polymerase. K.L.J. gave technical assistance. R.K.P. and K.L. contributed to manuscript preparation and enabled ChIP-seq and RNA-seq. A.G.-S. supervised and discussed the work with infectious influenza viruses. A.T. conceived and supervised this study and wrote the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.M. (imarazzi@rockefeller.edu) or A.T. (tarakho@rockefeller.edu).

Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling

Christoph Reinhardt^{1,2,3}, Mattias Bergentall^{1,2}, Thomas U. Greiner^{1,2}, Florence Schaffner⁴, Gunnel Östergren-Lundén^{1,2}, Lars C. Petersen⁵, Wolfram Ruf⁴ & Fredrik Bäckhed^{1,2,6}

The gut microbiota is a complex ecosystem that has coevolved with host physiology. Colonization of germ-free (GF) mice with a microbiota promotes increased vessel density in the small intestine¹, but little is known about the mechanisms involved. Tissue factor (TF) is the membrane receptor that initiates the extrinsic coagulation pathway², and it promotes developmental and tumour angiogenesis^{3,4}. Here we show that the gut microbiota promotes TF glycosylation associated with localization of TF on the cell surface, the activation of coagulation proteases, and phosphorylation of the TF cytoplasmic domain in the small intestine. Anti-TF treatment of colonized GF mice decreased microbiota-induced vascular remodelling and expression of the proangiogenic factor angiopoietin-1 (Ang-1) in the small intestine. Mice with a genetic deletion of the TF cytoplasmic domain or with hypomorphic TF (*F3*) alleles had a decreased intestinal vessel density. Coagulation proteases downstream of TF activate protease-activated receptor (PAR) signalling implicated in angiogenesis⁵. Vessel density and phosphorylation of the cytoplasmic domain of TF were decreased in small intestine from PAR1-deficient (*F2r*^{-/-}) but not PAR2-deficient (*F2rII*^{-/-}) mice, and inhibition of thrombin showed that thrombin-PAR1 signalling was upstream of TF phosphorylation. Thus, the microbiota-induced extravascular TF-PAR1 signalling loop is a novel pathway that may be modulated to influence vascular remodelling in the small intestine.

The mammalian intestine is an organ with marked postnatal vascular adaptation, which is induced at weaning and coincides with the development of an adult microbiota. In agreement with early studies showing that the gut microbiota affects vascular remodelling in the intestine¹, we showed significant increases in villus width in the small intestine of conventionally raised (CONV-R) mice in comparison with GF mice (Fig. 1a), suggesting a link between vascular remodelling and altered villus architecture on colonization. We also showed increased staining and messenger RNA levels of the vascular marker platelet-endothelial cell adhesion molecule 1 (PECAM-1) in the small intestine of both CONV-R and conventionalized (CONV-D; GF mice that had been colonized for 14 days with a normal microbiota from a CONV-R mouse) mice in comparison with GF mice (Fig. 1b-d). The increased vessel density was located to the mid-distal part of the small intestine (Supplementary Fig. 1). Staining for the tip-cell marker delta-like ligand 4 (Dll4)⁶ indicated that colonization initially promoted sprouting angiogenesis but that the number of tip cells returned to basal levels once villus remodelling was complete (Fig. 1e).

We found increased levels of mRNA for Ang-1 as well as increased phosphorylation of the Ang-1 receptor Tie-2 in the small intestine of CONV-R in comparison with GF mice (Fig. 1f,g), thus providing a potential mechanism for microbiota-induced vascular remodelling. Consistent with increased vessel density, vascular endothelial growth factor receptor 1 (VEGFR-1) expression was also higher in CONV-R mice, but there were no changes in any other components of the VEGF

pathway (Supplementary Fig. 2). The Ang-1-Tie-2 axis promotes the remodelling and sprouting of blood vessels^{7,8}. To confirm a role for Ang-1 in microbiota-induced vascular remodelling, we injected GF mice with the specific Ang-1 inhibitor mL4-3 before and during a 14-day colonization with a normal gut microbiota and showed decreases in Tie-2 phosphorylation and intestinal vessel density (Fig. 1h-j). We identified the epithelium as a source of Ang-1, because its expression was increased in isolated primary enterocytes from CONV-R mice in comparison with those from GF mice (Fig. 1k).

Angiogenesis is linked to the cellular initiation of coagulation, and TF signalling has been shown to modulate angiogenesis^{3,4}. Because bacterial components are known to stimulate the coagulation system⁹, we speculated that TF could have a function in microbiota-induced angiogenesis in the intestine. In agreement with earlier studies of TF localization in humans¹⁰ and mice¹¹, we identified TF predominantly in enterocytes of the villi of small intestine in both GF and CONV-R mice (Supplementary Fig. 3). We injected GF mice with anti-TF antibody or control IgG before and at 4 and 9 days after colonization with a normal caecal microbiota. Tissues were harvested 14 days after colonization, and we confirmed that the injected antibodies localized to the small intestine (Supplementary Fig. 4). Anti-TF treatment did not affect PECAM-1 staining in GF mice that were not colonized (Fig. 1l, m); neither were levels of VEGF-A, VEGFR-2 or VEGFR-3 mRNA in CONV-D mice affected (Supplementary Fig. 5). However, anti-TF treatment decreased villus width (Fig. 1n), vessel density (Fig. 1o, p) and expression of PECAM-1 and Ang-1 mRNA (Fig. 1q, r) in CONV-D mice, suggesting that TF promotes microbiota-induced remodelling of the villus vasculature. Similarly, the vessel density was decreased in intestines from mice expressing low levels of human TF (low-TF mice)¹² compared with mice expressing normal levels of human TF from a knocked-in minigene¹³ (Supplementary Fig. 6). Because neither humanized mouse strain expressed alternatively spliced TF, intestinal vascular remodelling seems to be independent of alternatively spliced TF¹⁴.

Paneth cells have been suggested to regulate microbiota-induced intestinal angiogenesis in mice, but they also have a large effect on angiogenesis independently of colonization status¹. Anti-TF treatment did not decrease the number of Paneth cells or mRNA levels of Paneth cell-derived cryptdin 2 in CONV-D mice (Supplementary Fig. 7), indicating that treatment with antibody has no cytotoxic effect on Paneth cells. In addition, vessel density was similar in colonized CR2-tox176 transgenic mice, which lack Paneth cells¹⁵, and their wild-type littermates after treatment with anti-TF (Supplementary Fig. 8).

Next we investigated whether intestinal TF expression and activity differed between GF and CONV-R mice. We did not observe any differences in intestinal levels of mRNA for TF from the two groups of mice (Fig. 2a). In contrast, immunoblot analyses identified two TF-reactive bands, one with an apparent molecular mass of 33 kDa that

¹Sahlgrenska Center for Cardiovascular and Metabolic Research/Wallenberg Laboratory, University of Gothenburg, 413 45 Gothenburg, Sweden. ²Department of Molecular and Clinical Medicine, University of Gothenburg, 413 45 Gothenburg, Sweden. ³Center for Thrombosis and Hemostasis (CTH), University Medical Center Mainz, Experimental Research, 55131 Mainz, Germany. ⁴Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California 92037, USA. ⁵Haemostasis Biology, Novo Nordisk A/S, DK-2760 Maalov, Denmark. ⁶Novo Nordisk Foundation Center for Basic Metabolic Research, Section for Metabolic Receptology and Enteroendocrinology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.

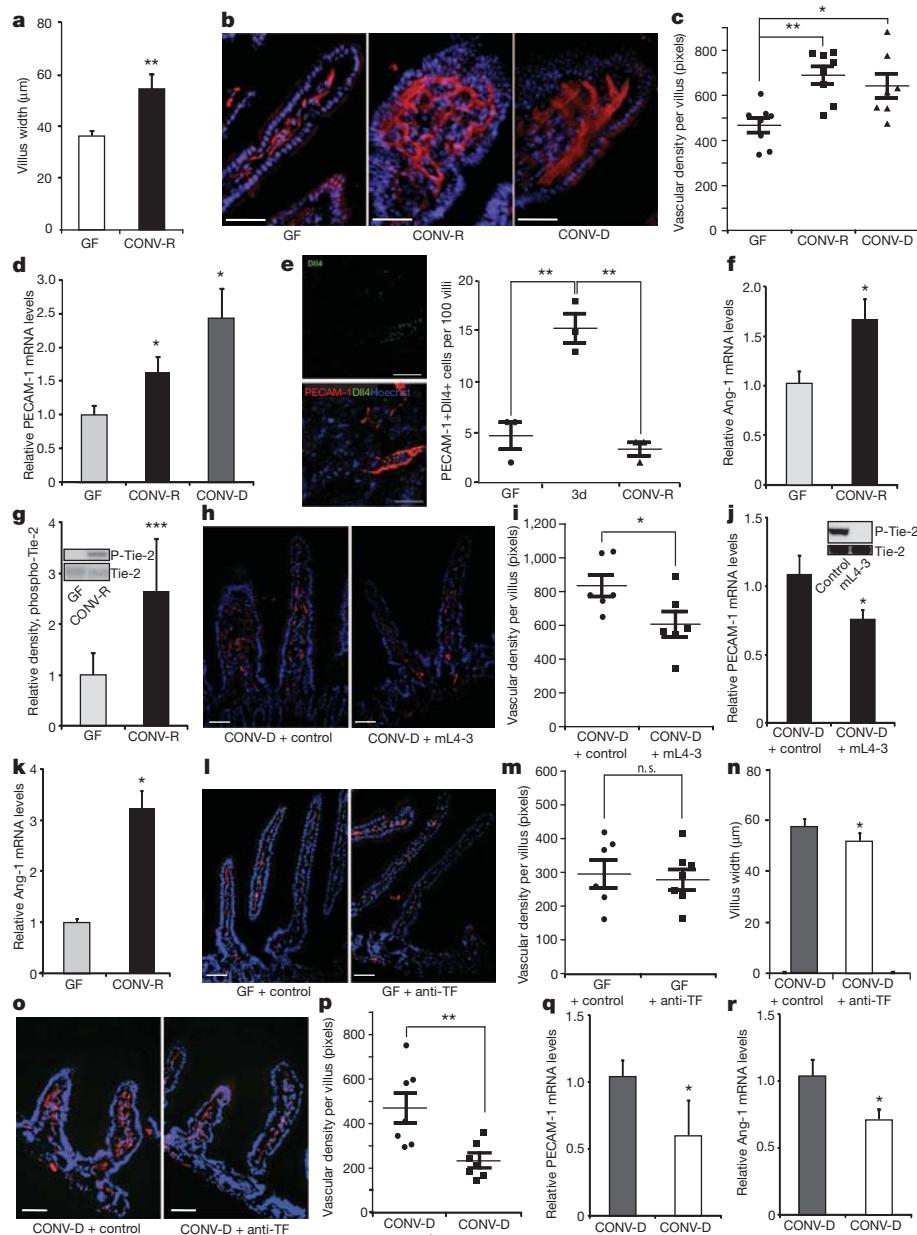


Figure 1 | TF promotes microbe-induced vascular remodelling in the gut.

a, Villus width of sections of small intestine from GF and CONV-R mice ($n = 4$ mice per group). **b**, PECAM-1 staining (red) of sections of small intestine from GF, CONV-R and CONV-D mice. Nuclei were stained with Hoechst nuclear dye (blue). **c**, Quantification of **b** ($n = 7$ or 8 mice per group). **d**, Relative levels of mRNA for the vascular marker PECAM-1 in GF, CONV-R and CONV-D mice ($n = 6$ or 7 mice per group). **e**, Dll4 staining (green) of sections of small intestine from GF, ex-GF mice colonized for 3 days (3d), and CONV-R mice. Endothelial cells were stained with PECAM-1 (red). Dll4-positive endothelial cells per 100 villi were quantified ($n = 3$ mice per group). **f**, Relative levels of mRNA for Ang-1 in sections of small intestine from GF and CONV-R mice ($n = 7$ –11 mice per group). **g**, Anti-phospho-Tie-2 immunoblot (Y1100 phosphorylation site) and quantification relative to total Tie-2 of small-intestinal lysates from GF and CONV-R mice ($n = 5$ mice per group). **h**, PECAM-1 staining of sections of small intestine from mice treated with control NaCl solution or the Ang-1 neutralizing peptibody mL4-3. **i**, Quantification of **h** ($n = 6$ mice per group). **j**, Relative levels of mRNA for

was present in both groups and a second band with an apparent molecular mass of 46 kDa that was present at higher levels in intestinal lysates from CONV-R and CONV-D mice (Fig. 2b, c). The gut microbiota has global effects on protein glycosylation in the small intestine¹⁶, which is necessary for the correct cellular localization and function of

PECAM-1 in sections of small intestine from CONV-D mice treated with NaCl control or mL4-3 ($n = 10$ or 11 mice per group). The inset shows that mL4-3 is a potent inhibitor of Ang-1-mediated Tie-2 phosphorylation. **k**, Relative levels of mRNA for Ang-1 in primary enterocytes from GF and CONV-R mice ($n = 10$ or 11 mice per group). **l**, PECAM-1 staining of sections of small intestine from GF mice treated with control or anti-TF antibody.

m, Quantification of **l** ($n = 6$ or 7 mice per group). **n**, Villus width of sections of small intestine from CONV-D mice treated with control or anti-TF antibody ($n = 4$ mice per group). **o**, PECAM-1 staining of sections of small intestine from CONV-D mice treated with control or anti-TF antibody.

p, Quantification of **o** ($n = 7$ mice per group). **q, r**, Relative levels of mRNA for PECAM-1 (**q**) and Ang-1 (**r**) in small intestine from CONV-D mice treated with control or anti-TF antibody ($n = 5$ or 6 mice per group). Female Swiss Webster mice or cells isolated from these mice were analysed in all panels. Scale bars, 50 μ m. Results are shown as means \pm s.e.m. Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.005$; n.s., not significant.

many proteins including TF procoagulant activity¹⁷. We therefore speculated that the 46-kDa TF band resulted from microbiota-induced N-linked glycosylation of TF, the primary carbohydrate modification of TF¹⁷. In agreement with this, the mannose-binding lectin concanavalin A readily detected the 46-kDa form of TF in

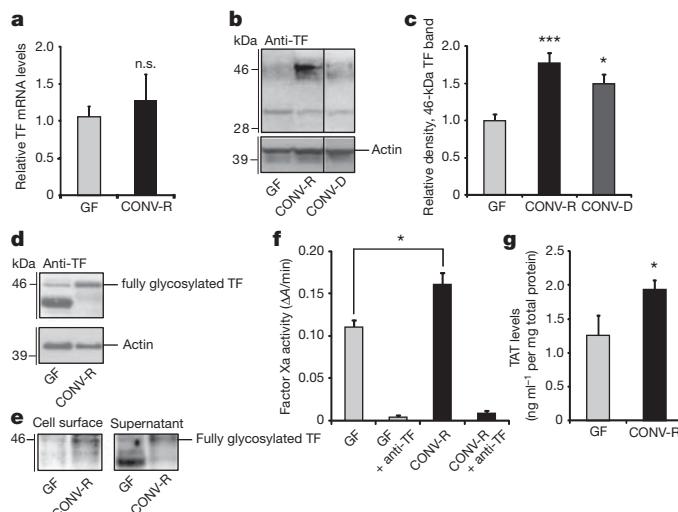


Figure 2 | The gut microbiota increases TF procoagulant activity and cell-surface localization. **a**, Relative levels of mRNA for TF in sections of small intestine from GF and CONV-R mice ($n = 7$ –11 mice per group). **b**, Anti-TF immunoblot of small-intestinal lysates from GF, CONV-R and CONV-D mice. **c**, Quantification of the 46-kDa TF band shown in **b** ($n = 14$ –25 mice per group). Data are normalized to actin and expressed relative to GF. **d**, Anti-TF immunoblot of primary enterocytes (from GF and CONV-R mice) after 2 h of culture. **e**, Anti-TF immunoblots from *N*-hydroxysuccinimidobiotin-labelled primary enterocytes from GF and CONV-R mice. Left: pull-down of proteins located on the plasma membrane with NeutrAvidin beads. Right: supernatant containing unlabelled proteins. **f**, Factor Xa activity in small-intestinal lysates from GF and CONV-R mice treated with control or anti-TF antibody (1H1; $n = 4$ or 5 mice per group). **g**, Levels of thrombin–antithrombin (TAT) complexes in small-intestinal lysates from GF and CONV-R mice ($n = 7$ mice per group). Female Swiss Webster mice or cells isolated from these mice were analysed in all panels. Results are shown as means \pm s.e.m. Asterisk, $P < 0.05$; three asterisks, $P < 0.005$; n.s., not significant.

small-intestinal lysates from CONV-D mice (Supplementary Fig. 9a). Treatment with the *N*-glycosidase PNGase F abolished detection of the 46-kDa form and generated a partly deglycosylated form with increased electrophoretic mobility that was only weakly detected by concanavalin A. We also treated primary enterocytes from CONV-R mice with the *N*-glycosylation inhibitor tunicamycin and observed a decreased abundance of the 46-kDa form (Supplementary Fig. 9b, c). These findings indicate that the gut microbiota promotes *N*-glycosylation of TF.

Exposure of functional TF on cell surfaces is regulated by basolateral sorting in epithelial cells¹⁸. Surface biotinylation followed by biotin pull-down of primary enterocytes from GF mice showed that the underglycosylated TF was mainly intracellular (Fig. 2d, e). In contrast, surface labelling of proteins or carbohydrates showed that enterocytes from CONV-R mice had high levels of the fully glycosylated TF on the cell surface (Fig. 2d, e and Supplementary Fig. 10). Confocal microscopy confirmed plasma membrane localization of TF in primary enterocytes isolated from CONV-R mice (Supplementary Fig. 11 and Supplementary Movie). These changes were associated with enhanced coagulation activation, as demonstrated by increased TF-VIIa-dependent generation of coagulation factor Xa and higher levels of thrombin–antithrombin complexes in lysates of small intestine from CONV-R in comparison with those from GF mice (Fig. 2f, g).

Not only does TF initiate coagulation, it also interacts with integrins on the extracellular side and regulates integrin function through its cytoplasmic domain¹⁹. Proximity ligation and immunoprecipitation experiments showed increased TF– β_1 integrin complex formation in intestinal tissue from CONV-R mice in comparison with GF counterparts (Supplementary Fig. 12a, b). Furthermore, TF– β_1 integrin complex formation was decreased by treating CONV-R mice with

tunicamycin (Supplementary Fig. 12c). The cytoplasmic domain of TF contains a conserved Ser/Thr-Pro phosphorylation site²⁰. Phosphorylation of this domain has been observed at sites of neovascularization⁴; it requires surface localization of TF²¹ and regulates integrin function¹⁹. An antibody directed against the phosphorylated domain of mouse TF detected increased phosphorylation in the 46-kDa form of TF in lysates of small intestine from CONV-R in comparison with that from GF mice (Fig. 3a), and treatment of primary enterocytes from CONV-R mice with tunicamycin decreased levels of the phosphorylated 46-kDa form of TF (Fig. 3b, c).

To test directly whether the TF cytoplasmic domain was involved in vascular remodelling, we analysed small-intestinal tissue from mice with a targeted deletion of this domain (Δ CT mice)²² and age-matched wild-type mice. Δ CT mice had significantly decreased villus vascularization (Fig. 3d, e) and decreased expression of mRNA for PECAM-1 and Ang-1 in comparison with wild-type mice (Fig. 3f). TF from wild-type and Δ CT mice had similar electrophoretic mobilities (Supplementary Fig. 13). These data show that the TF cytoplasmic domain has a function in increasing vessel density in the small intestine but that it is not required for glycosylation. Anti-TF treatment decreased TF phosphorylation but not total TF levels in CONV-D mice (Supplementary Fig. 14). These results indicate that the inhibitory effects of anti-TF on vascular remodelling are independent of TF downregulation but, at least in part, involve inhibition of TF cytoplasmic domain phosphorylation.

TF also mediates signalling through coagulation proteases that activate the G-protein-coupled receptors PAR1 and PAR2 (refs 4, 5). We investigated the effect of the gut microbiota on PAR expression in small-intestinal tissue. Levels of mRNA for PAR1 but not those for PAR2 were increased in CONV-R mice in comparison with GF

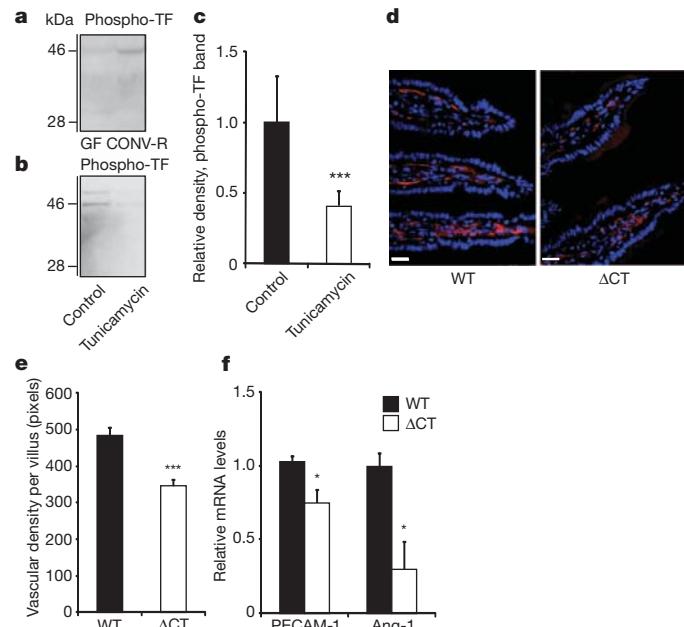


Figure 3 | The gut microbiota increases phosphorylation of the cytoplasmic tail of TF, which increases vessel density in the intestine. **a, b**, Anti-phospho-TF immunoblot of small-intestinal lysates from GF and CONV-R mice (**a**) and primary enterocytes (from CONV-R mice) incubated for 2 h in the absence and presence of tunicamycin ($10 \mu\text{mol l}^{-1}$) (**b**). **c**, Quantification of **b** ($n = 5$ mice per group). **d**, PECAM-1 staining (red) of sections of small intestine from 10–12-week-old wild-type (WT) and Δ CT female mice on a C57Bl6/J genetic background. Nuclei were stained with Hoechst nuclear dye (blue). **e**, Quantification of **d** ($n = 4$ –6 mice per group). **f**, Relative levels of mRNA for PECAM-1 and Ang-1 in segments of small intestine from WT and Δ CT mice ($n = 3$ or 4 mice per group). Female Swiss Webster mice or cells isolated from these mice were analysed in **a**–**c**. Scale bars, 20 μm . Results are shown as means \pm s.e.m. Asterisk, $P < 0.05$; three asterisks, $P < 0.005$.

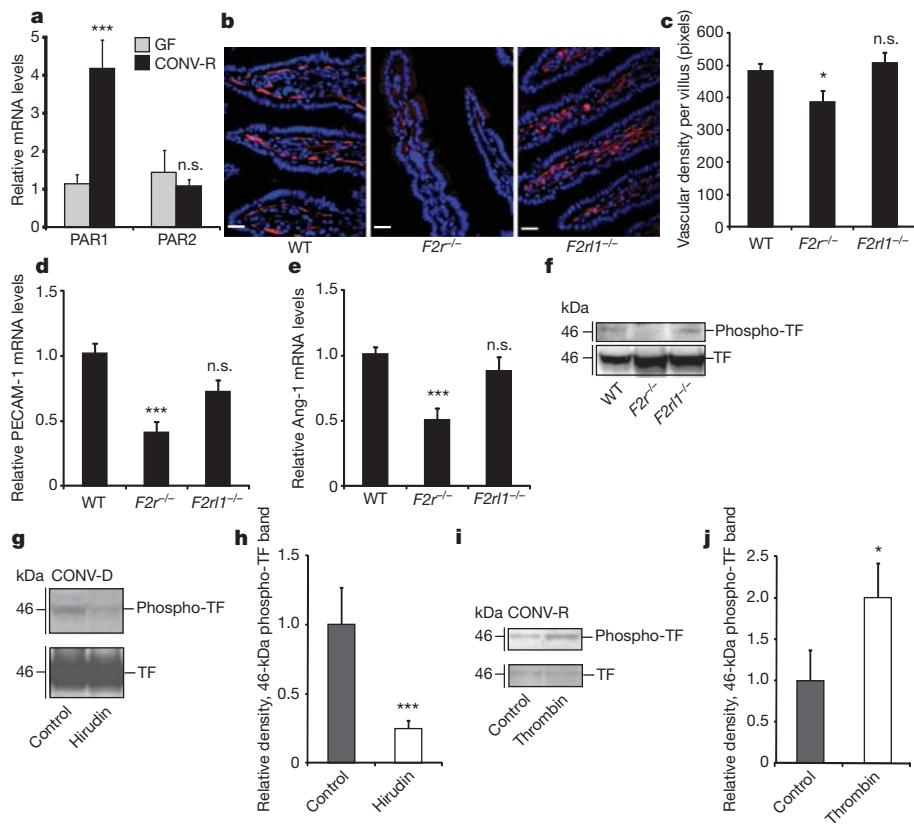


Figure 4 | PAR1 activation increases vessel density in the small intestine. **a**, Relative levels of mRNA for PAR1 and PAR2 in segments of small intestine from GF and CONV-R mice ($n = 7$ or 8 mice per group). **b**, PECAM-1 staining (red) of sections of small intestine from wild-type (WT), $F2r^{-/-}$ and $F2rl1^{-/-}$ mice. Nuclei were stained with Hoechst nuclear dye (blue). **c**, Quantification of **b** ($n = 6$ –9 mice per group). **d**, **e**, Relative levels of mRNA for PECAM-1 (**d**) and Ang-1 (**e**) in segments of small intestine from wild-type, $F2r^{-/-}$ and $F2rl1^{-/-}$ mice ($n = 6$ –9 mice per group). **f**, **g**, Anti-TF and anti-phospho-TF immunoblots of small-intestinal lysates from WT, $F2r^{-/-}$ and $F2rl1^{-/-}$ mice (**f**) and CONV-D mice treated with PBS (control) or hirudin (1 mg/mouse)

counterparts (Fig. 4a). PAR1 is abundantly expressed in endothelial cells⁵, but we found that PAR1 was also expressed in enterocytes and at higher levels in cells from CONV-R mice than in GF counterparts (Supplementary Fig. 15). PECAM-1 staining (Fig. 4b, c) as well as levels of mRNA for PECAM-1 and Ang-1 (Fig. 4d, e) were decreased in intestinal tissue from PAR1-deficient ($F2r^{-/-}$) but not PAR2-deficient ($F2rl1^{-/-}$) mice, which is in agreement with a study showing that thrombin induces PAR1-dependent Ang-1 expression in endothelial cells²³. Together, these data show that the microbiota induces increased expression of PAR1, and that PAR1 has a role in remodelling the vasculature in the small intestine.

We next investigated the potential interrelation between PAR1 and TF in intestinal tissue. Phosphorylation of TF was decreased in lysates of small intestine from $F2r^{-/-}$ in comparison with that from wild-type and $F2rl1^{-/-}$ mice (Fig. 4f), indicating that PAR1 acts upstream of TF phosphorylation. We blocked thrombin and thrombin-dependent PAR1 signalling with hirudin immediately before and during colonization of GF mice for 6 h, and observed a striking decrease in TF phosphorylation in lysates of small intestine (Fig. 4g, h). We also showed that thrombin increased the phosphorylation of TF in primary enterocytes (Fig. 4i, j). Taken together, these data suggest that functional, procoagulant TF is required for the generation of thrombin, which in turn activates PAR1 to promote phosphorylation of the cytoplasmic domain of TF in enterocytes.

This study has uncovered a novel connection between TF, PAR1 and Ang-1 in modulating vascular remodelling after colonization. Our

immediately before colonization and at 2 h and 4 h after colonization (g). **h**, Quantification of the phospho-TF band shown in **g** ($n = 6$ or 7 mice per group). **i**, Anti-TF and anti-phospho-TF immunoblots of primary enterocytes (from CONV-R mice) incubated for 2 h with human thrombin (50 nmol l⁻¹). **j**, Quantification of the phospho-TF band shown in **i** ($n = 8$ mice per group). Female Swiss Webster mice were analysed in **a** and **g–j**. Female WT, $F2r^{-/-}$ and $F2rl1^{-/-}$ mice on a C57BL/6J genetic background were used in **b–f**. Scale bars, 20 μ m. Results are shown as means \pm s.e.m. Asterisk, $P < 0.05$; three asterisks, $P < 0.005$; n.s., not significant.

results support a model in which the microbiota induces increased glycosylation and surface localization of TF in the small intestine, leading to activation of coagulation, PAR1-dependent phosphorylation of the TF cytoplasmic domain, and TF cytoplasmic domain signalling linked to Ang-1-dependent vascular remodelling (Supplementary Fig. 16). This pathway is distinct from established models of ocular angiogenesis⁴ or tumour-induced neovascularization, which requires the TF–Factor VIIa–PAR2-mediated induction of pro-angiogenic chemokines²⁴. We therefore suggest that TF may support distinct pro-angiogenic pathways in different tissues. Increased vascularization of the villi of the small intestine increases oxygenation of the villi, which are shortened and widened after colonization. This process may promote increased nutrient absorption, which has been associated with increased adiposity in CONV-R mice²⁵. Further dissection of how TF and PAR1 mediate postnatal microbiota-induced angiogenesis may provide new therapeutic targets for improving intestinal homeostasis and modulating the absorptive capacity of the gut.

METHODS SUMMARY

Mice. GF Swiss Webster female mice were maintained in flexible film isolators under a 12-h light cycle and fed with an autoclaved chow diet (Labdiet, St Louis) *ad libitum*. CONV-R Swiss Webster female mice were transferred into identical isolators at weaning. Age-matched female Δ CT (ref. 22), $F2r^{-/-}$ (ref. 26) and $F2rl1^{-/-}$ (ref. 27) mice and wild-type controls on a C57BL/6J background, low-TF and TF knock-in mice on a C57BL/6 background¹², and transgenic CR2-tox176 and non-transgenic littermates on a FVB/N genetic background¹⁵ were also used.

Mice were killed at 10–14 weeks of age by cervical dislocation or overdose anaesthesia; small intestines were removed and divided into eight equal segments. The fifth segment was used unless otherwise stated. Animal protocols were approved by the Research Animal Ethics Committee in Gothenburg and the Scripps Research Institute Institutional Animal Care and Use Committee (IACUC).

Isolation of primary enterocytes. Enterocytes were isolated from the fifth segment of small intestine as described previously²⁸. The cells were cultured for 2 h in DMEM medium before the beginning of the experiment.

Administration of TF antibody. Rabbit anti-mouse TF antibody²⁹ or rabbit anti-mouse IgG (Sigma) (1.33 mg per kg body weight) were administered intraperitoneally to GF mice before conventionalization with a caecal microbiota from a CONV-R donor. Additional antibody injections were given 4 and 9 days after colonization. The mice were killed 14 days after colonization.

Statistics. Data were analysed with Student's *t*-test for two sample groups or one-way analysis of variance for three sample groups.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 December 2009; accepted 23 January 2012.

Published online 11 March 2012.

1. Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc. Natl Acad. Sci. USA* **99**, 15451–15455 (2002).
2. Morrissey, J. H., Fakhrai, H. & Edgington, T. S. Molecular cloning of the cDNA for tissue factor, the cellular receptor for the initiation of the coagulation protease cascade. *Cell* **50**, 129–135 (1987).
3. Carmeliet, P. *et al.* Role of tissue factor in embryonic blood vessel development. *Nature* **383**, 73–75 (1996).
4. Belting, M. *et al.* Regulation of angiogenesis by tissue factor cytoplasmic domain signaling. *Nature Med.* **10**, 502–509 (2004).
5. Griffin, T. C., Srinivasan, Y., Zheng, Y.-W., Huang, W. & Coughlin, S. R. A role for thrombin receptor signaling in endothelial cells during embryonic development. *Science* **293**, 1666–1670 (2001).
6. Hellström, M. *et al.* DLL4 signalling through Notch 1 regulates formation of tip cells during angiogenesis. *Nature* **445**, 776–780 (2007).
7. Sato, T. N. *et al.* Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature* **376**, 70–74 (1995).
8. Suri, C. *et al.* Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell* **87**, 1171–1180 (1996).
9. Iwanaga, S. The limulus clotting reaction. *Curr. Opin. Immunol.* **5**, 74–82 (1993).
10. More, L. *et al.* Immunohistochemical study of tissue factor expression in normal intestine and idiopathic inflammatory bowel disease. *J. Clin. Pathol.* **46**, 703–708 (1993).
11. Luther, T. *et al.* Tissue factor expression during human and mouse development. *Am. J. Pathol.* **149**, 101–113 (1996).
12. Parry, G. C., Erlich, J. H., Carmeliet, P., Luther, T. & Mackman, N. Low levels of tissue factor are compatible with development and hemostasis in mice. *J. Clin. Invest.* **101**, 560–569 (1998).
13. Snyder, L. A. *et al.* Expression of human tissue factor under the control of the mouse tissue factor promoter mediates normal hemostasis in knock-in mice. *J. Thromb. Haemost.* **6**, 306–314 (2008).
14. van den Berg, Y. W. *et al.* Alternatively spliced tissue factor induces angiogenesis through integrin ligation. *Proc. Natl Acad. Sci. USA* **106**, 19497–19502 (2009).
15. Garabedian, E. M., Roberts, L. J., McNevin, M. S. & Gordon, J. I. Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. *J. Biol. Chem.* **272**, 23729–23740 (1997).
16. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host–microbial interactions in an open mammalian ecosystem. *Science* **273**, 1380–1383 (1996).
17. Krudysz-Amblo, J., Jennings, M. E. II, Mann, K. G. & Butenas, S. Carbohydrates and activity of natural and recombinant tissue factor. *J. Biol. Chem.* **285**, 3371–3382 (2010).
18. Camerer, E. *et al.* Opposite sorting of tissue factor in human umbilical vein endothelial cells and Madin–Darby canine kidney epithelial cells. *Blood* **88**, 1339–1349 (1996).
19. Dorfleutner, A., Hintermann, E., Tarui, T., Takada, Y. & Ruf, W. Cross-talk of integrin α 3 β 1 and tissue factor in cell migration. *Mol. Biol. Cell* **15**, 4416–4425 (2004).
20. Zioncheck, T. F., Sounmitra, R. & Vehar, G. A. The cytoplasmic domain of tissue factor is phosphorylated by a protein kinase C-dependent mechanism. *J. Biol. Chem.* **267**, 3561–3564 (1992).
21. Dorfleutner, A. & Ruf, W. Regulation of tissue factor cytoplasmic domain phosphorylation by palmitoylation. *Blood* **102**, 3998–4005 (2003).
22. Melis, E. *et al.* Targeted deletion of the cytoplasmic domain of tissue factor in mice does not affect development. *Biochem. Biophys. Res. Commun.* **286**, 580–586 (2001).
23. Blackburn, J. S. & Brinckerhoff, C. E. Matrix metalloproteinase-1 and thrombin differentially activate gene expression in endothelial cells via PAR-1 and promote angiogenesis. *Am. J. Pathol.* **173**, 1736–1746 (2008).
24. Schaffner, F. *et al.* Cooperation of tissue factor cytoplasmic domain and PAR2 signaling in breast cancer development. *Blood* **116**, 6106–6113 (2010).
25. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
26. Connolly, A. J., Ishihara, H., Kahn, M. L., Farese, R. V. Jr & Coughlin, S. R. Role of the thrombin receptor in development and evidence for a second receptor. *Nature* **381**, 516–519 (1996).
27. Damiano, B. P. *et al.* Cardiovascular responses mediated by protease-activated receptor-2 (PAR-2) and thrombin receptor (PAR-1) are distinguished in mice deficient in PAR-2 or PAR-1. *J. Pharmacol. Exp. Ther.* **288**, 671–678 (1999).
28. Perreault, N. & Beaulieu, J.-F. Primary cultures of fully differentiated and pure human intestinal epithelial cells. *Exp. Cell Res.* **245**, 34–42 (1998).
29. Furlan-Freguia, C., Marchese, P., Gruber, A., Ruggeri, Z. M. & Ruf, W. P2X7 receptor signaling contributes to tissue factor-dependent thrombosis in mice. *J. Clin. Invest.* **121**, 2932–2944 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Perkins for editing the manuscript; C. Arvidsson, A. Hallén, S. Wagoner, M. Karlsson, D. O'Donell, S. Islam, N. Hörmann and A. Mohammadzadeh for technical assistance; A. Hallén for providing Supplementary Fig. 16; and P. Lindahl, J. Gordon, C. Betsholtz, M. Bergö, A. Wichmann, V. Tremaroli, M. Levin and S. Massberg for comments and suggestions. We are grateful to D. Kirchhofer for the gift of 1H1 monoclonal anti-mouse TF antibody, J. Nichols at Amgen for mL4-3, N. Mackman for the low-TF mice, M. Anderson for the human TF knock-in mice, and J. Gordon for providing CR2-tox176 mice. This study was supported by the Swedish Foundation for Strategic Research, the Swedish Research Council, Torsten and Ragnar Söderberg's foundation, Petrus and Augusta Hedlund's foundation, and the Swedish federal government under the LUA/ALF agreement to F.B., National Institutes of Health grants HL-60742 and HL-77753 to W.R., and a Marie Curie Fellowship, a Marie Curie Reintegration Grant from the European Union and the German Federal Ministry of Education and Research (BMBF 01EO1003) to C.R.

Author Contributions C.R. was responsible for conception and study design, biochemical analysis of TF, analysis of vessel densities, data assembly and analysis, and writing the manuscript. M.B., T.U.G., F.S. and G.Ö.L. performed data collection, analysis and interpretation and commented on the manuscript. L.C.P. provided material. W.R. and F.B. were responsible for conception and study design, data analysis and interpretation, and writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.B. (frederik.backhed@wlab.gu.se).

METHODS

Administration of mL4-3. mL4-33 (ref. 30) (2.32 mg per kg body weight) was administered subcutaneously to GF mice before conventionalization with a caecal microbiota from a CONV-R donor. Additional injections of mL4-3 were given three times a week. The mice were killed 14 days after colonization.

Preparation of intestinal samples. For immunohistochemistry and *in situ* hybridization, the small intestine (divided into eight equal segments) and colon were flushed with PBS after excision and opened longitudinally. The tissue was fixed overnight in 4% formaldehyde at 4 °C, washed three times in PBS, and incubated in 10% sucrose in PBS at 4 °C. After 3 h, the buffer was replaced with 20% sucrose and 10% glycerol in PBS, and the tissue was incubated at 4 °C overnight. Tissues were dried with a paper towel and mounted in OCT on solid CO₂. Frozen sections 6 µm thick were prepared.

For mRNA analyses, the segments were frozen immediately at -80 °C in liquid nitrogen. For immunoblots, the fifth segment was flash-frozen and homogenized for 10 min in lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100) containing Roche Complete protease and PhosStop phosphatase inhibitors (diluted 1:10). The homogenate was incubated for 30 min on ice and centrifuged three times at 9,000g for 10 min to remove insoluble cell debris.

Immunohistochemistry. Sections were incubated for 20 min at 22 °C and blocked for 1 h with diluted TBST (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% Triton X-100) containing 5% rabbit serum. The blocking solution was removed and the following primary antibodies, diluted in the same blocking solution, were added: rat anti-mouse PECAM-1 (dilution 1:300; BD, Franklin Lakes), chicken anti-cytokeratin 8 (dilution 1:100; Abcam, Cambridge), goat anti-DLL4 (dilution 1:50; R&D), rabbit anti-mouse TF²⁹ (1 µg ml⁻¹) and rabbit anti-PAR1 (dilution 1:300; Sigma). The samples were incubated overnight at 4 °C, washed three times for 5 min in TBST and incubated for 1 h with secondary antibodies (Invitrogen, Carlsbad) at room temperature (rabbit anti-rat Alexa594, dilution 1:800; goat anti-rabbit IgG Alexa488, dilution 1:5,000; goat anti-chicken IgG Alexa488, dilution 1:2,000; all from BD). Nuclei were stained with Hoechst dye (3 µg ml⁻¹; Sigma) and the sections were washed three times for 10 min in TBST. For detection of Paneth cells, fluorescein isothiocyanate-isolectin (10 µg ml⁻¹; Sigma) was used. Slides were mounted, and viewed at $\times 20$ and $\times 40$ magnification with a fluorescence microscope (Axioplan 2 imaging; Zeiss, Oberkochen). Biopix iQ software (<http://www.biopix.se>) was used to quantify PECAM-1 staining in 2–11 villi per mouse. Confocal images and three-dimensional reconstructions were obtained with a Leica TCS SP5 confocal microscope (Leica, Wetzlar).

Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analysis. Total RNA was isolated from small-intestinal tissues and isolated primary enterocytes with the RNeasy kit (Qiagen, Hilden). Total RNA (0.5 µg) was reverse transcribed (High Capacity cDNA Reverse Transcription kit; Applied Biosystems, Foster City) and SYBR green-based qRT-PCR was performed as described previously³¹. Primers are listed in Supplementary Table 1.

In situ hybridization. Mouse TF cDNA³² was subcloned into pSPT19 for subsequent *in vitro* RNA synthesis. Non-radioactive, digoxigenin-labelled sense and antisense RNA probes were synthesized with the DIG RNA Labelling Kit (SP6/7; Roche, Mannheim). Tissues were pretreated for 2 min with proteinase K (10 µg ml⁻¹) in 50 mM Tris-HCl pH 7.5, 5 mM EDTA; the reaction was stopped by washing for 30 s in 0.2% glycine in PBS, followed by two additional washing steps in PBS. Tissues were fixed for 15 min in 4% paraformaldehyde in PBS, and washed in PBS for 2 min. Hybridization solution was added, and tissues were pre-hybridized for 1 h at 65 °C. RNA probe (8 ng µl⁻¹ hybridization solution) was added and preheated for 5 min at 80 °C; 100 µl was added to each slide and incubated overnight at 65 °C in a humidified box. Slides were washed three times for 30 min in a preheated washing solution at 65 °C and twice for 30 min in MABT (100 mM maleic acid pH 7.5, 150 mM NaCl, 0.1% Tween 20) at room temperature. Slides were blocked with 2% blocking reagent (Roche, Mannheim), 20% heat-inactivated sheep serum (Sigma) in MABT for 1 h at room temperature. Binding of the RNA template was detected with alkaline-phosphatase-conjugated Fab fragments (Roche, Mannheim) and BM Purple.

Factor Xa activity. Factor Xa activity was measured in small-intestinal lysates as described previously³².

Measurement of TAT complexes. The TAT ELISA Kit (Uscnlife, Guangguoguji) was used for determination of the concentration of mouse TAT complexes in lysates of small-intestinal tissue.

Immunoprecipitation. Tissue lysates were incubated for 1 h with anti-mouse TF antibody (70 µg ml⁻¹; American Diagnostica, Stamford) or anti-integrin β₁ antibody (dilution 1:100; Cell Signaling, Danvers), and immunocomplexes were precipitated by adding 50 µl of Protein A-Sepharose fast flow 4B (Sigma). TF and integrin β₁ antigen were detected as described below.

Glycosidase treatment. Anti-TF precipitates were boiled for 5 min to release the captured antigen from the antibody. Samples were cooled to 4 °C, and 20 U ml⁻¹ peptide N-glycosidase F (Sigma) was added for 90 min at 37 °C and then boiled again for 5 min to inactivate the glycosidase. Treatment with O-glycosidase (25 mU ml⁻¹; Merck, Darmstadt) was performed for 3 h at 37 °C.

Immunoblotting. Tissue lysates or immunoprecipitates were separated by using a NuPAGE system with MOPS buffer and 10% BisTris gels. Proteins were transferred to poly(vinylidene difluoride) membranes (Invitrogen, Carlsbad). The membrane was blocked in 5% milk powder (in PBS/Tween) and incubated for 1.5 h in 5% milk powder containing the primary antibody (rabbit anti-mouse TF (2.5 µg ml⁻¹; American Diagnostica, Stamford) for immunoprecipitation, rabbit anti-mouse TF and rabbit anti-mouse phospho-TF (2 µg ml⁻¹) for specificity controls—see Supplementary Fig. 17, rabbit anti-integrin β₁ (dilution 1:1,000; Cell Signaling), rabbit anti-actin (dilution 1:200; Sigma), rabbit anti-phospho-Tie2 (dilution 1:250; R&D) and rabbit anti-Tie-2 antibody (dilution 1:250; Abcam)). Secondary goat anti-rabbit IgG (horseradish peroxidase-conjugated; Santa Cruz Biotechnology, Santa Cruz) was applied for 1 h. Alternatively, the membrane was incubated with horseradish peroxidase-conjugated concanavalin A (Sigma) to detect sugar moieties after immunoprecipitation. Then the membrane was first washed for 2 min with PBS and incubated overnight with the lectin solution (PBS containing Mg²⁺ and Ca²⁺). The next day, the blot was rinsed three times with PBS/Tween. Blots were developed with enhanced chemiluminescence solutions (Amersham Biosciences, Little Chalfont). For densitometric analysis of protein bands, the software Multi Gauge V3.0 (Fuji Film, Tokyo) was applied.

Cell-surface labelling and pull-down with N-hydroxysuccinimido-biotin. For amine-reactive biotinylation and isolation of cell surface proteins from isolated primary enterocytes, the Cell Surface Protein Isolation Kit (Pierce, Rockford) was used. Isolated proteins were separated on a 10% BisTris gel (Invitrogen), and TF antigen was analysed by immunoblotting.

Proximity ligation assay³³. Slides with adhering primary enterocytes were blocked and incubated with primary antibodies (monoclonal rat anti-mouse TF (1H1)³⁴, 23.4 µg ml⁻¹, provided by Daniel Kirchhofer; rabbit polyclonal anti-integrin β₁, dilution 1:50; Cell Signaling Technology). Secondary antibodies (anti-rat and anti-rabbit) conjugated with unique DNA probes (Olink Bioscience, Uppsala) were added. Slides were evaluated with a Leica TCS SP5 confocal microscope. If TF and integrin β₁ antigens are closer than 30 nm, a fluorescence signal can be generated.

30. Falcón, B. L. *et al.* Contrasting actions of selective inhibitors of angiopoietin-1 and angiopoietin-2 on the normalization of tumor blood vessels. *Am. J. Pathol.* **175**, 2159–2170 (2009).
31. Stappenbeck, T. S. *et al.* Laser capture microdissection of mouse intestine: characterizing mRNA and protein expression, and profiling intermediary metabolism in specified cell populations. *Methods Enzymol.* **356**, 167–196 (2002).
32. Petersen, L. C. *et al.* Characterization of recombinant murine factor VIIa and recombinant murine tissue factor: a human–murine species compatibility study. *Thromb. Res.* **116**, 75–85 (2005).
33. Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes *in situ* by proximity ligation. *Nature Methods* **3**, 995–1000 (2006).
34. Kirchhofer, D., Moran, P., Bullens, S. & Peale, F. A monoclonal antibody that inhibits mouse tissue factor function. *J. Thromb. Haemost.* **3**, 1098–1099 (2005).

The *Shigella flexneri* effector OspI deamidates UBC13 to dampen the inflammatory response

Takahito Sanada¹, Minsoo Kim¹, Hitomi Mimuro², Masato Suzuki³, Michinaga Ogawa³, Akiho Oyama³, Hiroshi Ashida³, Taira Kobayashi³, Tomohiro Koyama⁴, Shinya Nagai⁴, Yuri Shibata⁵, Jin Gohda⁵, Jun-ichiro Inoue⁵, Tsunehiro Mizushima⁶ & Chihiro Sasakawa^{1,3,4}

Many bacterial pathogens can enter various host cells and then survive intracellularly, transiently evade humoral immunity, and further disseminate to other cells and tissues. When bacteria enter host cells and replicate intracellularly, the host cells sense the invading bacteria as damage-associated molecular patterns (DAMPs) and pathogen-associated molecular patterns (PAMPs) by way of various pattern recognition receptors. As a result, the host cells induce alarm signals that activate the innate immune system¹. Therefore, bacteria must modulate host inflammatory signalling and dampen these alarm signals^{2–4}. How pathogens do this after invading epithelial cells remains unclear, however. Here we show that OspI, a *Shigella flexneri* effector encoded by *ORF169b* on the large plasmid and delivered by the type III secretion system, dampens acute inflammatory responses during bacterial invasion by suppressing the tumour-necrosis factor (TNF)-receptor-associated factor 6 (TRAF6)-mediated signalling pathway. OspI is a glutamine deamidase that selectively deamidates the glutamine residue at position 100 in UBC13 to a glutamic acid residue. Consequently, the E2 ubiquitin-conjugating activity required for TRAF6 activation is inhibited, allowing *S. flexneri* OspI to modulate the diacylglycerol-CBM (CARD–BCL10–MALT1) complex–TRAF6–nuclear-factor- κ B signalling pathway. We determined the 2.0 Å crystal structure of OspI, which contains a putative cysteine–histidine–aspartic acid catalytic triad. A mutational analysis showed this catalytic triad to be essential for the deamidation of UBC13. Our results suggest that *S. flexneri* inhibits acute inflammatory responses in the initial stage of infection by targeting the UBC13–TRAF6 complex.

The rupture of host cell membranes by invasive bacteria such as *S. flexneri* and *Listeria monocytogenes* is sensed as DAMPs and triggers acute inflammatory responses by activating various cell signals^{5,6}. Within the cytoplasm, *S. flexneri* multiplies and subsequently spreads to neighbouring cells and, during this dissemination, the bacteria release peptidoglycan, lipopolysaccharide and flagellin^{3,7}. These bacterial components are recognized as PAMPs by cytoplasmic pattern recognition receptors, such as NOD-like receptors (NLRs), which induce nuclear factor- κ B (NF- κ B)-mediated and mitogen-activated protein kinase (MAPK)-mediated inflammatory responses³. To counteract this host defence, bacteria have evolved mechanisms to modulate host inflammatory responses³, through delivering several effectors that modulate inflammatory signalling by way of the type III secretion system (T3SS)^{8,9}. To better understand these mechanisms, we searched for additional *S. flexneri* effectors that modulate acute inflammatory responses to bacterial invasion, and we found that OspI has a pivotal role (Supplementary Fig. 1).

To assess the role of OspI, we carried out a comprehensive microarray analysis of HeLa cells infected with each of three *S. flexneri* strains: YSH6000 (wild type), Δ ospI (*ospI*-deficient) or S325 (T3SS-deficient).

We found that the messenger RNA levels of chemokines (for example, interleukin-8 (IL-8), CC-chemokine ligand 20 (CCL20), CXC-chemokine ligand 2 (CXCL2) and CCL2) and cytokines (tumour-necrosis- α (TNF- α) and IL-6) were greatly increased in Δ ospI-infected cells at 60 min after infection (Supplementary Fig. 2a). This elevated chemokine and cytokine production was detected as early as 30 min after infection (Fig. 1a and Supplementary Fig. 2b). Increased phosphorylation of inhibitor of NF- κ B (IkB α) was detected in Δ ospI-infected HeLa cells relative to YSH6000-infected HeLa cells as early as 10 min after infection (Fig. 1b). The nuclear translocation of NF- κ B (p65 subunit) was fourfold higher in cells infected with Δ ospI compared with YSH6000 at 20 min after infection (Fig. 1c and Supplementary Fig. 3), suggesting that OspI can dampen the acute inflammatory response.

We next tested whether OspI inhibits NF- κ B activation when *S. flexneri* infects cells, and we found that ectopic OspI expression further inhibited NF- κ B activation on *S. flexneri* infection (Fig. 1d). When *S. flexneri* infects epithelial cells, NOD1–RIP2 (nucleotide-binding oligomerization domain 1–receptor-interacting serine/threonine kinase 2)-dependent and NOD1–RIP2-independent NF- κ B signalling pathways are both activated. PAMPs stimulate the dependent pathway¹⁰, whereas DAMPs stimulate the independent pathway⁶. Therefore, we investigated the levels of NF- κ B activation in HeLa cells transiently expressing OspI (designated HeLa/OspI cells) or a mock control. When HeLa/OspI cells were stimulated with TNF- α , NOD1 or phorbol myristate acetate (PMA), an examination of the levels of activated NF- κ B revealed that OspI suppressed PMA-mediated NF- κ B activation but not TNF α - or NOD1-mediated NF- κ B activation (Fig. 1d).

Because PMA is a substitute for diacylglycerol (DAG) in the activation of the protein kinase C (PKC)–NF- κ B pathway¹¹, and because DAG in the host membrane can trigger antibacterial autophagy against *Salmonella enterica* serovar Typhimurium¹², we examined the membrane ruffles protruding around *S. flexneri* entry sites in cells expressing PKC-C1–GFP (the PKC-C1 region fused to green fluorescent protein (GFP) as a DAG sensor)¹³. We found that DAG accumulated around the bacterial entry site (Supplementary Fig. 4a). Indeed, the increased *IL8* mRNA production in Δ ospI-infected cells was suppressed by treating the cells with propranolol, a DAG inhibitor (Supplementary Fig. 4b).

The DAG–NF- κ B pathway is mediated through the CBM complex, a major regulator of NF- κ B signalling in lymphoid, myeloid and non-myeloid cells in innate and adaptive immune responses¹⁴. We examined whether OspI modulates CBM-complex-mediated NF- κ B signalling, through knocking down *BCL10* production with shorting interfering RNA (siRNA), and we found that the *IL8* mRNA levels were greatly decreased compared with cells treated with control siRNA (Fig. 1e and Supplementary Fig. 4c). GFP–MALT1 was recruited to the *S. flexneri* entry point in HeLa/GFP–MALT1 cells because MALT1 functionally

¹Department of Infectious Disease Control, International Research Center for Infectious Diseases, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ²Division of Bacteriology, Department of Infectious Disease Control, International Research Center for Infectious Diseases, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ³Division of Bacterial Infection, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ⁴Nippon Institute for Biological Science, 9-2221-1 Shinmachi, Ome, Tokyo 198-0024, Japan. ⁵Division of Cellular and Molecular Biology, Department of Cancer Biology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ⁶Picobiology Institute, Department of Life Science, Graduate School of Life Science, University of Hyogo, 3-2-1 Kouto, Kamigori, Akoh, Hyogo 678-1297, Japan.

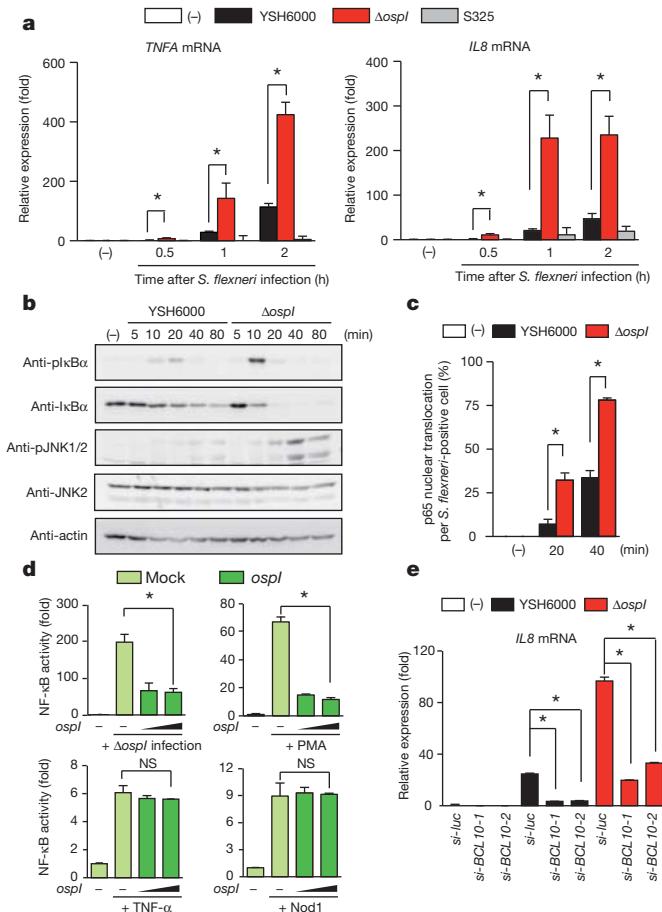


Figure 1 | OspI inhibits DAG-mediated NF-κB activation during *S. flexneri* infection. **a**, The *TNFA* and *IL8* expression levels in HeLa cells at 0.5, 1 and 2 h post infection with YSH6000, Δ ospI or S325. The white bar (–) indicates no infection. **b**, Phosphorylation of I κ B α and JNK1/2 in HeLa cells infected with YSH6000 or Δ ospI at various times after infection, as detected with the indicated antibodies. pI κ B α , phosphorylated I κ B α ; pJNK1/2, phosphorylated JNK1/2. **c**, Nuclear localization of p65 in HeLa cells infected with YSH6000 or Δ ospI at 20 and 40 min post infection. **d**, The luciferase activity of NF-κB at 90 min post infection is shown. NS, not significant. **e**, The *IL8* expression levels at 60 min post infection in HeLa cells transfected with *BCL10*-directed siRNA and infected with YSH6000 or Δ ospI. Data are presented as the mean \pm s.d.; $n = 3$; *, $P < 0.05$. *si-luc*, siRNA-luciferase.

interacts with BCL10 (Supplementary Fig. 4d). These results suggested that *S. flexneri*-induced changes in the host cell membrane stimulate the DAG-CBM-complex–NF-κB pathway and that OspI specifically dampens this pathway.

To gain further structural and functional insight, we determined the crystal structure of recombinant *S. flexneri* OspI at 2.0 Å resolution (Supplementary Table 1; Protein Data Bank (PDB) ID 3B21). OspI has an α / β fold with four β -strands (β 1– β 4), seven α -helices (α 1– α 7) and one β 10 helix (Fig. 2a, b).

A search of known structures in PDB revealed that OspI shares structural homology with a cysteine protease family and is most closely related to AvrPphB (Supplementary Fig. 5). AvrPphB is a *Pseudomonas syringae* effector that is delivered by way of a T3SS and has been classified into a superfamily of enzymes containing cysteine proteases, acetyl transferases, deamidases, transglutaminases and ubiquitin carboxy-terminal hydrolase^{15,16}. Although there is considerable divergence in the overall folding across this superfamily, a core anti-parallel β -sheet and an α -helix that contains the active site cysteine, which packs against the β -strands, are conserved across this family (Supplementary Fig. 6). A potential catalytic triad (C62, H145 and D160) in OspI was identified based on a comparison with the active site of

AvrPphB (Fig. 2c and Supplementary Figs 5 and 6a). Although overall amino acid sequence similarity was low between OspI and these proteins, superimposing H145 and D160 of OspI onto H212 and D227 of AvrPphB or other members of this superfamily revealed remarkable similarity^{17–21} (Fig. 2c).

C62 of OspI, however, existed in three discrete conformations in the crystal structure, and the S γ position was located on the opposite side of the active site in AvrPphB. The fractional occupancy of each conformer was estimated to be 0.55 (conformation A), 0.35 (conformation B) and 0.1 (conformation C). The highest occupancy site of C62 appeared to form a disulphide bond with C65 at 2.05 Å (Fig. 2c, d and Supplementary Fig. 7). Alternative conformations of a deamidase active site cysteine are also seen in cytotoxic necrotizing factor 1 (Cnf1)²².

To characterize the putative catalytic triad, we substituted C62, H145 and D160 with serine (C62S) or alanine (C62A, H145A and D160A), and the mutant constructs were tested for their ability to suppress NF-κB activation. Complementing the Δ ospI mutant with plasmids carrying the *ospI*-C62S, *ospI*-H145A or *ospI*-D160A genes failed to mitigate the increased I κ B α phosphorylation and *IL8* induction induced by Δ ospI infection (Fig. 3a, b and Supplementary Fig. 8c). OspI-C62S, OspI-H145A and OspI-D160A lost the ability to suppress NF-κB activity on *S. flexneri* infection or PMA stimulation (Supplementary Fig. 8a, b), indicating that C62, H145 and D160 in OspI compose the catalytic triad that suppresses NF-κB signalling.

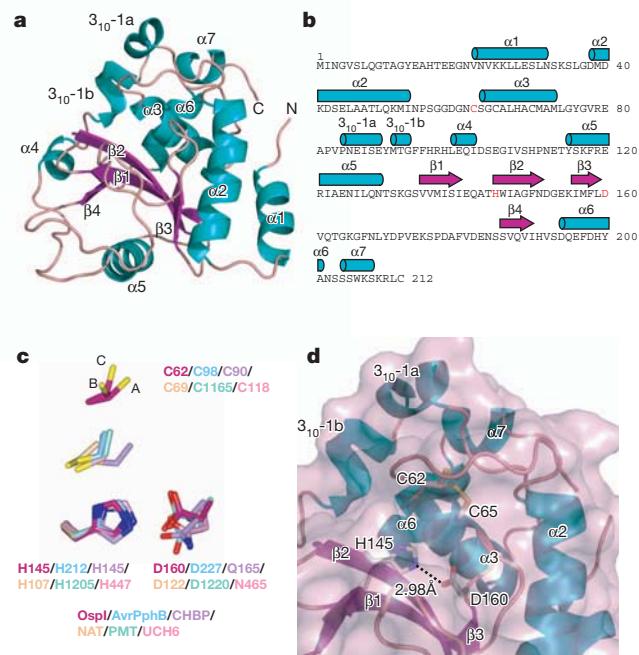
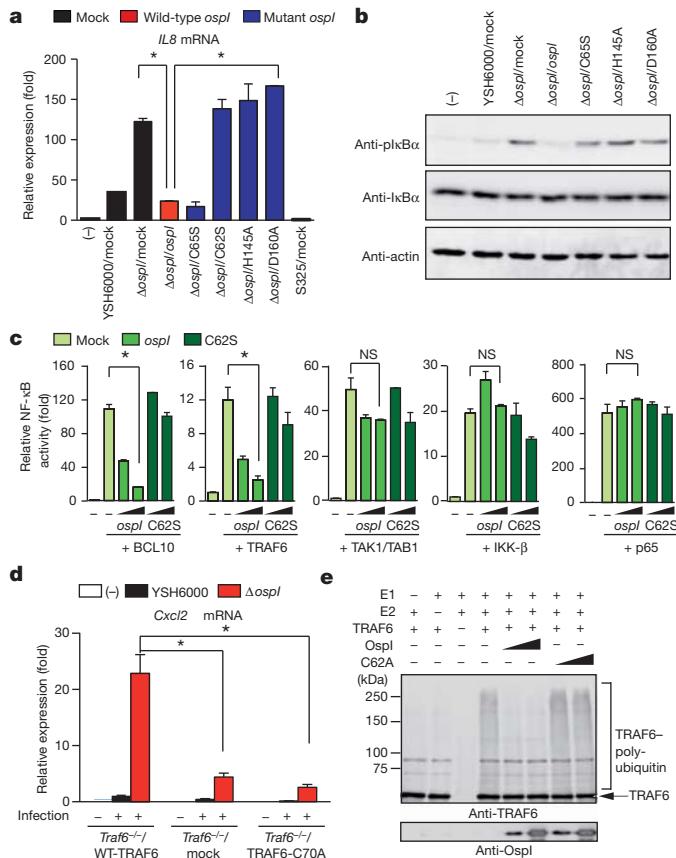


Figure 2 | Crystal structure of *S. flexneri* OspI. **a**, Overall structure of *S. flexneri* OspI. The secondary structure elements are coloured as follows: α -helices, cyan; β -strands, dark pink; and loops, light pink. **b**, The amino acid sequence of OspI. Secondary structure elements are coloured as in **a**. The putative active site residues are shown in red. **c**, Alignments of the catalytic cores of OspI with *P. syringae* AvrPphB and other members of the superfamily to which AvrPphB belongs, using the putative catalytic triad residues. All atoms of histidine and the main chain atoms of aspartic acid are shown as a reference. Shown are *S. flexneri* OspI, AvrPphB (PDB ID 1UKF), CHBP (Cif homologue from *Burkholderia pseudomallei*; PDB ID 3GQM), NAT (arylamine N-acetyltransferase from *Salmonella Typhimurium*; PDB ID 1E2T), PMT (*Pasteurella multocida* toxin; PDB ID 2EBF), and UCH6 (ubiquitin C-terminal hydrolase 6; PDB ID 1VJV). C62 in OspI is represented in three alternative conformations with the three conformers labelled A, B and C. **d**, A close-up view of the putative active site of OspI. OspI is shown with a surface representation. The putative active site residues are shown as stick models. Hydrogen bonding is indicated by a dashed line.



We determined which steps in the DAG-CBM-complex-NF-κB pathway are targeted by OspI by examining NF-κB activity in HeLa/OspI or HeLa/OspI-C62S cells that were transfected with a vector expressing the gene encoding BCL10, TRAF6, TAK1/TAB1, IKK-β or NF-κB (p65). OspI but not OspI-C62S suppressed NF-κB activity when HeLa cells were transfected with the gene encoding BCL10 or TRAF6 but not TAK1/TAB1, IKK-β or p65, suggesting that OspI targets TRAF6 or an upstream step in the signalling pathway (Fig. 3c). Thus, we tested rescued *Traf6*-deficient mouse embryonic fibroblasts (*Traf6*^{-/-} MEFs) for a possible role of OspI in modulating TRAF6 activation during *S. flexneri* infection: the MEFs tested were *Traf6*^{-/-}/WT-TRAF6 (*Traf6*^{-/-} MEFs expressing wild-type human TRAF6), *Traf6*^{-/-}/TRAF6-C70A (an E3-ligase-deficient mutant) and *Traf6*^{-/-}/mock. *Cxcl2* levels in Δ ospI-infected *Traf6*-deficient MEFs were low and could be rescued by wild-type TRAF, but not by TRAF6-C70A or mock treatment (Fig. 3d and Supplementary Fig. 8d). This result is consistent with OspI interference in TRAF6 activation during *S. flexneri* infection.

TRAF6 is an E3 ubiquitin ligase that cooperates with ubiquitin-activating E1 and ubiquitin-conjugating E2 enzymes (UBC13 and UEV1A), which are required for TRAF6 self-ubiquitylation and TRAF6-induced NF-κB activation^{23–26}. Therefore, we investigated the effects of OspI and OspI-C62A on TRAF6 by using a self-ubiquitylation assay. We found that OspI, but not OspI-C62A, dampedened TRAF6 polyubiquitylation (Fig. 3e and Supplementary Fig. 9). However, OspI did not affect the formation of E2-ubiquitin thioester intermediates

(Supplementary Fig. 10), suggesting that OspI modifies TRAF6, UBC13, UEV1A or ubiquitin. We incubated OspI with each of these putative targets and examined their electrophoretic mobility. We found that the mobility of UBC13, but not of the other proteins, shifted in an OspI-dose-dependent manner (Supplementary Fig. 11). Indeed, OspI could interact with His-UBC13 (Supplementary Fig. 12).

We then identified two affected peptides (WSPALQIR and DKWSPALQIR, which overlap each other) from a tryptic digest of UBC13 by using tandem mass spectrometry, and we found that the Q100 residue in UBC13 was deamidated to E100 by OspI (Fig. 4a and Supplementary Fig. 13). To confirm this, we created UBC13-Q100E and showed that this modification produced the same mobility shift as the modification at Q100 by OspI; treatment with OspI-C62A, OspI-H145A or OspI-D160A produced no mobility shift (Fig. 4b, c and Supplementary Fig. 14). Endogenous UBC13 was consistently deamidated in HeLa cells 10 min after infection with YSH6000 and Δ ospI/ospI but not Δ ospI or Δ ospI/ospI-C62S (Fig. 4d).

Importantly, the cocrystal structures of UBC13 and the zinc finger of TRAF6 indicated that Q100 of UBC13 was near the catalytic pocket but was also located near the interface between UBC13 and the TRAF6 zinc finger²⁷. Thus, we further characterized the ubiquitin-conjugating E2 activity of UBC13-Q100E using a ubiquitylation assay. The efficiency of UBC13-Q100E-catalysed ubiquitin chain formation on TRAF6 was markedly lower than that of wild-type UBC13 (Fig. 4e). In an NF-κB reporter assay, UBC13-Q100E acted as a dominant negative because UBC13-Q100E suppressed the NF-κB activity that was stimulated by PMA, TRAF6 and infection, but not TNF-α, in a dose-dependent manner (Supplementary Fig. 15). We thus concluded that OspI deamidates Q100 in UBC13, inhibiting the TRAF6-NF-κB pathway.

Here we identified OspI as a new T3SS effector that specifically targets TRAF6-dependent acute inflammatory signalling during *S. flexneri* invasion of epithelial cells. OspI selectively deamidates UBC13, severely impairing the E2 ubiquitin ligase activity of UBC13, which is required for TRAF6 polyubiquitylation. Thus, *S. flexneri* can block acute NF-κB-mediated inflammatory responses at the early stage of epithelial invasion (Supplementary Fig. 1), suggesting that OspI is a unique T3SS effector

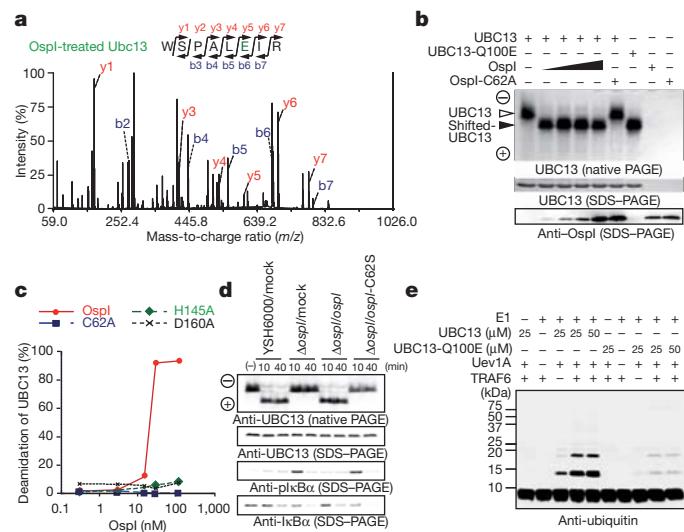


Figure 4 | OspI selectively deamidates Q100 in UBC13. **a**, Liquid chromatography and tandem mass spectrometry (LC-MS/MS) spectrum of a Q100-containing tryptic peptide in OspI-treated UBC13. **b**, Native PAGE and SDS-PAGE analysis of UBC13 treated with OspI (0.08, 0.8, 8 or 80 μ M) or OspI-C62A (80 μ M). **c**, Native PAGE analysis of UBC13 treated with OspI, OspI-C62A, OspI-H145A or OspI-D160A. **d**, HeLa cells were infected with *S. flexneri*/mock, Δ ospI/mock, Δ ospI/ospI or Δ ospI/ospI-C62S. At 10 and 40 min after infection, the cell lysates were subjected to native PAGE followed by anti-UBC13, anti-IκBα or anti-phospho-IκBα immunoblotting. **e**, *In vitro* ubiquitylation assay in the presence of UBC13 or UBC13-Q100E.

that dampens TRAF6-dependent inflammatory signalling in response to bacterial invasion of epithelial cells.

METHODS SUMMARY

The *S. flexneri* strains used in this study are described in the Methods. The procedures for the OspI crystallization and structure determination, mass spectrometric analysis, microarray analysis, quantitative PCR with reverse transcription, *in vitro* ubiquitylation assays, *in vitro* deamidation assays, luciferase assays, immunostaining, immunoblotting, *S. flexneri* infection and protein expression are described in the Methods, together with a description of the antibodies and the materials. OspI, reported as ORF169b (NCBI accession number CAC05849), is encoded by an *S. flexneri* virulence plasmid. The *S. flexneri* *ospI* gene was disrupted using a λ red recombinase-mediated recombination system as previously described²⁸. The *ΔospI* strain was complemented by introducing plasmids (pWKS130) encoding *ospI*-C62S-myc, *ospI*-H145A-myc or *ospI*-D160A-myc. Bacterial infections were carried out as previously described²⁹. Briefly, HeLa cells were infected with the various *S. flexneri* strains at a multiplicity of infection (m.o.i.) of 100. For the afimbral adhesin (Afa)-expressing *S. flexneri* strains, cells were infected at an m.o.i. of 10. Twenty minutes after infection, the plates were transferred into fresh medium containing 100 µg ml⁻¹ gentamicin, to kill extracellular bacteria. Statistical significance was determined using Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 July 2011; accepted 24 January 2012.

Published online 11 March 2012.

- Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
- Kim, M. et al. Bacterial interactions with the host epithelium. *Cell Host Microbe* **8**, 20–35 (2010).
- Ashida, H., Ogawa, M., Kim, M., Mimuro, H. & Sasakawa, C. Bacteria and host interactions in the gut epithelial barrier. *Nature Chem. Biol.* **8**, 36–45 (2011).
- Taxman, D. J., Huang, M. T. & Ting, J. P. Inflammasome inhibition as a pathogenic stealth mechanism. *Cell Host Microbe* **8**, 7–11 (2010).
- Hilbi, H. Bacterial jailbreak sounds cellular alarm: phagosome membrane remnants trigger signalling. *Cell Host Microbe* **6**, 102–104 (2009).
- Dupont, N. et al. *Shigella* phagocytic vacuolar membrane remnants participate in the cellular response to pathogen invasion and are regulated by autophagy. *Cell Host Microbe* **6**, 137–149 (2009).
- Ashida, H. et al. Cell death and infection: a double-edged sword for host and pathogen survival. *J. Cell Biol.* **195**, 931–942 (2011).
- Mattoo, S., Lee, Y. M. & Dixon, J. E. Interactions of bacterial effector proteins with host proteins. *Curr. Opin. Immunol.* **19**, 392–401 (2007).
- Rahman, M. M. & McFadden, G. Modulation of NF-κB signalling by microbial pathogens. *Nature Rev. Microbiol.* **9**, 291–306 (2011).
- Girardin, S. E. et al. CARD4/Nod1 mediates NF-κB and JNK activation by invasive *Shigella flexneri*. *EMBO Rep.* **2**, 736–742 (2001).
- Holden, N. S. et al. Phorbol ester-stimulated NF-κB-dependent transcription: roles for isoforms of novel protein kinase C. *Cell. Signal.* **20**, 1338–1348 (2008).
- Shahnazari, S. et al. A diacylglycerol-dependent signalling pathway contributes to regulation of antibacterial autophagy. *Cell Host Microbe* **8**, 137–146 (2010).
- Oancea, E., Teruel, M. N., Quest, A. F. & Meyer, T. Green fluorescent protein (GFP)-tagged cysteine-rich domains from protein kinase C as fluorescent indicators for diacylglycerol signalling in living cells. *J. Cell Biol.* **140**, 485–498 (1998).
- Rawlings, D. J., Sommer, K. & Moreno-Garcia, M. E. The CARMA1 signalosome links the signalling machinery of adaptive and innate immunity in lymphocytes. *Nature Rev. Immunol.* **6**, 799–812 (2006).
- Shao, F., Merritt, P. M., Bao, Z., Innes, R. W. & Dixon, J. E. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* **109**, 575–588 (2002).
- Zhu, M., Shao, F., Innes, R. W., Dixon, J. E. & Xu, Z. The crystal structure of *Pseudomonas* avirulence protein AvrPphB: a papain-like fold with a distinct substrate-binding site. *Proc. Natl Acad. Sci. USA* **101**, 302–307 (2004).
- Crow, A. et al. Crystal structures of Cif from bacterial pathogens *Photobacterium luminescens* and *Burkholderia pseudomallei*. *PLoS ONE* **4**, e5582 (2009).
- Sinclair, J. C., Sandy, J., Delgoda, R., Sim, E. & Noble, M. E. Structure of arylamine N-acetyltransferase reveals a catalytic triad. *Nature Struct. Biol.* **7**, 560–564 (2000).
- Kitadokoro, K. et al. Crystal structures reveal a thiol protease-like catalytic triad in the C-terminal region of *Pasteurella multocida* toxin. *Proc. Natl Acad. Sci. USA* **104**, 5139–5144 (2007).
- Orth, J. H. et al. *Pasteurella multocida* toxin activation of heterotrimeric G proteins by deamidation. *Proc. Natl Acad. Sci. USA* **106**, 7179–7184 (2009).
- Wang, Y. et al. MMDB: annotating protein sequences with Entrez's 3D-structure database. *Nucleic Acids Res.* **35**, D298–D300 (2007).
- Buetow, L., Flatau, G., Chiu, K., Boquet, P. & Ghosh, P. Structure of the Rho-activating domain of *Escherichia coli* cytotoxic necrotizing factor 1. *Nature Struct. Biol.* **8**, 584–588 (2001).
- Wang, C. et al. TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).
- Deng, L. et al. Activation of the IκB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000).
- Lamothe, B. et al. Site-specific Lys-63-linked tumor necrosis factor receptor-associated factor 6 auto-ubiquitination is a critical determinant of IκB kinase activation. *J. Biol. Chem.* **282**, 4102–4112 (2007).
- Fukushima, T. et al. Ubiquitin-conjugating enzyme Ubc13 is a critical component of TNF receptor-associated factor (TRAF)-mediated inflammatory responses. *Proc. Natl Acad. Sci. USA* **104**, 6371–6376 (2007).
- Yin, Q. et al. E2 interaction and dimerization in the crystal structure of TRAF6. *Nature Struct. Mol. Biol.* **16**, 658–666 (2009).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
- Ashida, H. et al. A bacterial E3 ubiquitin ligase IpaH9.8 targets NEMO/IKK γ to dampen the host NF-κB-mediated inflammatory response. *Nature Cell Biol.* **12**, 66–73 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Fukuda for matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) analysis. We thank the members of the Sasakawa laboratory for their advice. We are grateful to R. Whittier for critical reading of the manuscript. Diffraction data were collected at the Osaka University beamline BL44XU at SPring-8. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (23121525 (T.M.)), a Grant-in-Aid for Specially Promoted Research (23000012 (C.S.)), a Grant-in-Aid for Young Scientists (A) (23689027 (M.K.)), several Grants-in-Aid for Young Scientists (B) (23790471 (M.O.), 23790472 (H.A.) and 22790403 (T.S.)), a Grant-in-Aid for Scientific Research (B) (23390102 (H.M.)), a Grant-in-Aid for Challenging Exploratory Research (23659220 (H.M.)), a Grant-in-Aid for Scientific Research on Priority Areas (18073003 (C.S.)) and the Japan Initiative for Global Research Network on Infectious Diseases (C.S.). Part of this work was supported by grants from the Naito Foundation (M.K. and H.M.), the Waksman Foundation of Japan (M.O.), the Yakult Bio-Science Foundation (M.O.), the Yakult Central Institute (C.S.) and The Hayashi Memorial Foundation for Female Natural Scientists (M.K.).

Author Contributions T.S. and T.M. designed and performed the experiments. M.K., H.M., M.S., H.A., A.O., T. Kobayashi and M.O. assisted with the experiments. J.G., Y.S. and J.I.I. gave advice regarding the design of the experiments and provided TRAF6 materials. T. Koyama and S.N. made antibodies. A.O. made the *ΔospI* mutant. C.S. and T.M. wrote the paper.

Author Information Atomic coordinates and structure factors for the OspI structure have been deposited in the PDB under ID 3B21. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.S. (sasakawa@ims.u-tokyo.ac.jp) and T.M. (mizushi@sci.u-hyogo.ac.jp).

METHODS

Strains and plasmids. *S. flexneri* strain YSH6000 (ref. 30) was used as the wild-type strain. S325 (*mxiA*::Tn5)³¹ was used as a T3SS-deficient negative control. The vectors pCMV-FLAG-I κ B kinase- β (IKK- β), pCMV-FLAG—transforming growth factor- β -activated kinase 1 (TAK1), and pCMV-FLAG-TAK1-binding protein 1 (TAB1) were from G. Takaesi³². TRAF6 (ref. 33) was ligated into the plasmids pGEX6P-1, pCMV-FLAG and pMx-puro. Human *MALTI* and human *BCL10* cDNAs were ligated into the pCMV-FLAG plasmid. Site-directed mutagenesis of *ospI*, *UBC13* and *TRAF6* was performed using a QuikChange Site-Directed Mutagenesis Kit (Stratagene).

Antibodies and materials. Antiserum specific for OspI was obtained by immunizing rabbits with the *S. flexneri* OspI peptide (189–207 amino acids, VIHVSDQEFQFDHYANSSSWK) coupled to keyhole limpet haemocyanin using *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester. Anti-*S. flexneri* lipopolysaccharide was prepared as previously described³⁴. Anti-phospho-I κ B α (Cell Signaling Technology), anti-I κ B α (BD Transduction Laboratories), anti-phospho-JNK1/2 (Cell Signaling Technology), anti-JNK2 (Santa Cruz Biotechnology), anti-p65 (F-6; Santa Cruz Biotechnology), anti-M2 Flag (Sigma), anti-His (Sigma), anti-actin (Millipore), anti-UBC13 (Invitrogen) and anti-ubiquitin (1B3; MBL) antibodies, as well as Alexa Fluor 633 phalloidin (Invitrogen), were obtained commercially and used as primary antibodies for immunostaining and immunoblotting. TNF- α (Peprotech), PMA (Sigma) and propranolol (Sigma) were obtained commercially.

Plasmid construction, expression and protein purification for crystallization.

To purify OspI, the corresponding DNA was subcloned into the pGEX6P-1 glutathione S-transferase (GST) fusion vector (GE Healthcare). The plasmids were used to transform the BL21 strain of *Escherichia coli*, and protein expression was induced by adding 0.1 μ M isopropyl- β -D-thiogalactoside (IPTG) at 30 °C for 6 h. Whole-cell extracts were prepared by sonicating and incubating the bacteria in lysis buffer (20 mM Tris-HCl, pH 7.4, and 150 mM NaCl) for 30 min with rotation at 4 °C, and then clarified by centrifugation at 27,000g for 20 min. GST fusion proteins were purified by incubating the clarified cell extract with Glutathione Sepharose 4B. The GST moiety was removed by cleaving with the PreScission Plus protease (GE Healthcare) and then performing anion-exchange chromatography (HiTrap Q FF, GE Healthcare) and gel-filtration chromatography (Superdex 75, GE Healthcare). A selenomethionine (Se-Met) derivative of OspI was prepared by expressing the protein in strain B834 *E. coli* cells that were grown in minimal medium supplemented with L-selenomethionine (25 μ g ml⁻¹) (Calbiochem).

Crystallization. Native and Se-Met OspI were concentrated to 20 mg ml⁻¹ by ultrafiltration in 25 mM Tris-HCl, pH 7.5, and 1 mM dithiothreitol. Crystals were grown by the hanging-drop vapour diffusion method at 288 K in drops containing a mixture of 2 μ l of protein solution and 2 μ l of reservoir solution. For native and Se-Met OspI, the reservoir solution consisted of 0.1 M MES, pH 6.5, 0.1 M sodium acetate, 1.0 M NaCl and 24% (w/v) PEG 8000. The crystals of native and Se-Met OspI were transferred into a cryoprotective solution containing 0.1 M MES, pH 6.5, 0.1 M sodium acetate, 1.0 M NaCl and 35% (w/v) PEG 8000.

Data collection. X-ray diffraction data were collected at SPring-8 at 100 K in BL44XU. Single-wavelength anomalous diffraction (SAD) data were collected at a wavelength of 0.9789 Å. Native OspI crystal data were collected at a wavelength of 0.9000 Å. All data were reduced using the software Denzo, Scalepack³⁵ and programs from the CCP4 package³⁶. The data-processing statistics are provided in Supplementary Table 1.

Structure determination and refinement. The structure of OspI was determined using the SAD method. The positions of the heavy atoms were obtained using the software SHELXD³⁷ and refined using MLPHARE. The initial SAD phases were extended to a higher resolution with diffraction data that were collected from the native crystal to 2.0 Å resolution, with solvent flattening using PIRATE and DM³⁶. The initial model was constructed with the ARP/wARP³⁸ program. The remaining structural elements were built manually using the Coot³⁹ program. The model was refined to 2.0 Å resolution with the software REFMAC5 and CNS⁴⁰. The final OspI model shows amino acids 21 to 212. The phasing and refinement statistics are summarized in Supplementary Table 1. There are no residues in the disallowed regions of the Ramachandran plot. The structure figures were generated using CCP4mg⁴¹ and PyMOL⁴².

Cell culture and retroviral expression. HeLa and 293T cells were cultured in minimal essential medium (Sigma) and Dulbecco's modified Eagle's medium (Sigma), respectively, containing 10% FBS. *Traf6*-deficient MEFs were as described previously⁴³. Human TRAF6 and the C70A mutant were ligated into pMx-puro. The plasmids were transfected into Plat-E packaging cells as previously described³⁴. The resultant retrovirus was used to infect *Traf6*-deficient MEFs and selected for puromycin (2 μ g ml⁻¹) resistance.

Luciferase reporter assay. For the luciferase reporter assays, cells were transfected with the reporter plasmid (pGL4.32[*luc2P/NF- κ B-RE/Hygro*]; Promega), *Renilla* luciferase constructs (phRL-TK; Promega) and various combinations of expression

plasmids, by using FuGENE6 transfection reagent (Roche). The cells were infected with *S. flexneri* (m.o.i of 20) or treated with PMA (50 nM) and TNF- α (2 ng ml⁻¹) for 1.5 h.

siRNA knockdown analysis. HeLa cells were transfected with siRNA (RNAi Co) by reverse transfection using Lipofectamine RNAiMAX. The following siRNAs were used: human *BCL10* siRNA-1 (5'-CCUUAAGAUCACGUACGUUU-3' and 5'-ACAGUACGUGAUCUUAAGGG-3'); *BCL10* siRNA-2 (5'-CGUAC UGUUUCACGACAAUGA-3' and 5'-AUUGUCGUGAACAGUACGUG-3'); and *si-luc*, 5'-CGUACGCCGAAUACUUCGAdTdT-3' and 5'-UCGAAGAUU UCCGGCUACGdTdT-3'.

RNA extraction and quantitative RT-PCR analysis. Total RNA was extracted using ISOGEN (Nippongene). First-strand cDNA was synthesized from 1 μ g total RNA with reverse transcriptase using oligo(dT) primers. Real-time PCR was performed on cDNA samples using a LightCycler 2.0 (Roche) with the SYBR Green system (TaKaRa). The *GAPDH* expression levels were evaluated as an internal control. The following primer pairs were used: human *IL8* (5'-CTGA TTTCTGCAGCTCTGTGTG-3' and 5'-GTCCACTCTCAATCACTCTCAG-3'), human *TNF* (5'-CTTCAGACACCCCTCAACCTCTT-3' and 5'-CACATTCTT GAATCCCCAGGT-3'), human *CCL20* (5'-TTGATGTCAGTGTGCTACTCTC-3' and 5'-CCGTGTGAAGCCCACAATA-3'), human *CCL2* (5'-GCTCATAGCA GCCACCTTCATT-3' and 5'-CAGCTCTTTGGGACACTTGCT-3'), human *CXCL2* (5'-GGGTGGCAAAGAAAAGGAG-3' and 5'-GTTGAGCGTCAAGAC CCAGT-3'), human *IL6* (5'-GATGGCTGAAAAGATGGATGC-3' and 5'-CT GCAGGAACCTGGATCAGGACT-3'), human *GAPDH* (5'-TGCCCTCAACGA CCACTTTG-3' and 5'-TTCCTCTTGCTCTGCTGGG-3'), mouse *Cxcl2* (5'-CAAGGGTTGACTTCAAGAACATCC-3' and 5'-CCTTGAGAGTGGCTA TGACTTC-3'), mouse *Gapdh* (5'-GTGTCITTCACCACCATGGAG-3' and 5'-TCGTGGTTCACACCCATCAC-3') and human *BCL10* (5'-AGCGCAGAC ATCGGAGAGGT-3' and 5'-TGTGGCCGCAGAATGGCAGG-3').

Purification of recombinant proteins and in vitro ubiquitylation assay. E1 (Boston Biochem) and ubiquitin (Sigma) were obtained commercially. His-UEV1A, His-UBC13 and GST-TRAF6 were prepared as previously described⁴⁴. E1 (200 ng), His-UEV1A (400 ng), His-UBC13 (500 ng) and TRAF6 (300 ng) were incubated at 30 °C for 20 min in a 50 μ l reaction mixture containing reaction buffer (20 mM Tris-HCl, pH 7.5, 5 mM ATP, 5 mM MgCl₂ and 0.1 mM dithiothreitol (DTT)).

In vitro deamidation assay and mass spectrometric analysis. UBC13 or ubiquitin (12 μ g) was incubated with the indicated amount of purified OspI at 30 °C for 30 min in reaction buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl and 0.1 mM DTT). The reaction samples were separated by native PAGE, stained with Coomassie Brilliant Blue, and then quantified with the program ImageJ. The protein bands were excised from the gel, and each gel slice was subjected to in-gel digestion with trypsin. The resultant tryptic peptides were injected into a nano-liquid chromatography system (DiNa, KYA). The eluent was mixed with the matrix solution, and the mixture was directly blotted onto a matrix-assisted laser desorption (MALDI) sample plate with a micro-fractionation system (AccuSpot, Shimadzu Biotech) for MALDI-time of flight (MALDI-TOF) mass spectrometry (MS) analysis. Overall peptide identification was performed using a MALDI-TOF/TOF 4600 proteomics analyser (Applied Biosystems), followed by a database search with Mascot (Matrix Science).

Immunofluorescence analysis. HeLa cells were seeded on cover slips in six-well plates. The cells were infected with YSH6000 (wild-type *S. flexneri*) (m.o.i. of 100) for the indicated time periods, washed with PBS and then fixed in 4% paraformaldehyde in PBS for 15 min. The cells were permeabilized with 0.2% Triton X-100 in PBS for 10 min and then blocked with 1% BSA in TBS for 30 min.

Congo-red-induced secretion. Congo-red-induced secretion of type III effectors was performed as reported previously⁴⁵. Briefly, *S. flexneri* incubated at 37 °C for 4 h was stimulated with 0.003% Congo red for 10 min at 37 °C. The supernatant was recovered by centrifugation. The samples were separated by SDS-PAGE and immunoblotted with anti-MYC antibody.

Microarray analysis. Total RNA was prepared 1 h after infection and analysed by a gene microarray (GeneChip Human Genome U133 Plus 2.0 Array, Affymetrix). The raw data were normalized and analysed using GeneSpring software. Fold changes in the gene expression values (relative to uninfected cells) were log₂-transformed and visualized as a heat map, indicated by the colour scale.

- Sasakawa, C. et al. Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and Congo red binding activity in *Shigella flexneri*. *Infect. Immun.* **51**, 470–475 (1986).
- Sasakawa, C., Makino, S., Kamata, K. & Yoshikawa, M. Isolation, characterization, and mapping of Tn5 insertions into the 140-megadalton invasion plasmid defective in the mouse Sereny test in *Shigella flexneri* 2a. *Infect. Immun.* **54**, 32–36 (1986).

32. Takaesu, G. *et al.* TAB2, a novel adaptor protein, mediates activation of TAK1 MAPKKK by linking TAK1 to TRAF6 in the IL-1 signal transduction pathway. *Mol. Cell* **5**, 649–658 (2000).
33. Ishida, T. *et al.* Identification of TRAF6, a novel tumor necrosis factor receptor-associated factor protein that mediates signaling from an amino-terminal domain of the CD40 cytoplasmic region. *J. Biol. Chem.* **271**, 28745–28748 (1996).
34. Kim, M. *et al.* Bacteria hijack integrin-linked kinase to stabilize focal adhesions and block cell detachment. *Nature* **459**, 578–582 (2009).
35. Otwinowski, Z. M. W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
36. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
37. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
38. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244 (2003).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
41. McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D* **67**, 386–394 (2011).
42. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14–20 (2002).
43. Kobayashi, N. *et al.* Segregation of TRAF6-mediated signaling pathways clarifies its role in osteoclastogenesis. *EMBO J.* **20**, 1271–1280 (2001).
44. Morikawa, H. *et al.* The bacterial effector Cif interferes with SCF ubiquitin ligase function by inhibiting deneddylation of Cullin1. *Biochem. Biophys. Res. Commun.* **401**, 268–274 (2010).
45. Parsot, C., Menard, R., Gounon, P. & Sansonetti, P. J. Enhanced secretion through the *Shigella flexneri* Mxi-Spa translocon leads to assembly of extracellular proteins into macromolecular structures. *Mol. Microbiol.* **16**, 291–300 (1995).

Intrinsic coupling of lagging-strand synthesis to chromatin assembly

Duncan J. Smith¹ & Iestyn Whitehouse¹

Fifty per cent of the genome is discontinuously replicated on the lagging strand as Okazaki fragments. Eukaryotic Okazaki fragments remain poorly characterized and, because nucleosomes are rapidly deposited on nascent DNA, Okazaki fragment processing and nucleosome assembly potentially affect one another. Here we show that ligation-competent Okazaki fragments in *Saccharomyces cerevisiae* are sized according to the nucleosome repeat. Using deep sequencing, we demonstrate that ligation junctions preferentially occur near nucleosome midpoints rather than in internucleosomal linker regions. Disrupting chromatin assembly or lagging-strand polymerase processivity affects both the size and the distribution of Okazaki fragments, suggesting a role for nascent chromatin, assembled immediately after the passage of the replication fork, in the termination of Okazaki fragment synthesis. Our studies represent the first high-resolution analysis—to our knowledge—of eukaryotic Okazaki fragments *in vivo*, and reveal the interconnection between lagging-strand synthesis and chromatin assembly.

During eukaryotic chromosome replication, both genetic and epigenetic information must be accurately duplicated. For chromatin architecture and modifications to be truly epigenetic—that is, heritable despite not being genetically encoded—complete disruption and dissociation of nucleosomes from the replication fork must be prevented to allow the rapid re-deposition of precisely located and appropriately modified histones on the nascent DNA. Histone chaperone complexes govern both nucleosome disassembly and assembly at the replication fork¹.

DNA replication is inherently asymmetric. Okazaki fragment synthesis on the lagging strand necessitates the repeated production of single-stranded DNA and polymerization in the opposite direction to fork progression. Given the delay in lagging-strand synthesis and the rapidity of histone deposition behind the replication fork², these two processes may be interlinked. A coordinated series of events occurs each time an Okazaki fragment is synthesized³. Each fragment is initiated via an RNA primer⁴ and up to 30 nucleotides of DNA, both synthesized by DNA polymerase α (Pol α)-primase⁵; subsequently, the clamp loading factor RFC mediates the replacement of Pol α by the sliding clamp PCNA and the processive polymerase Pol δ (ref. 6). Pol δ extends the nascent DNA chain through the 5' end of the preceding Okazaki fragment^{7,8}; this strand-displacement synthesis generates a 5' RNA or DNA flap that is cleaved by nucleases such as the flap endonuclease Fen1 (ref. 9). Repeated cycles of extension and DNA cleavage followed by Pol δ idling¹⁰ produce a nick that migrates away from the replication fork and can be sealed by DNA ligase I¹¹. Little to no DNA synthesized by the error-prone Pol α remains in the genome after replication¹², suggesting that strand displacement generally replaces at least 30 nucleotides of DNA, but it remains unclear how the replication machinery coordinates the transition from Pol δ to DNA ligase. Indeed, despite their fundamental importance in replication, comparatively little is known about the *in vivo* properties of eukaryotic Okazaki fragments: although they are widely accepted to be shorter than the 1–2 kilobases (kb) observed in prokaryotes, a clear consensus has not emerged^{11,13–15}. Moreover, we have little information about how nucleosome assembly may interact with lagging-strand synthesis. A plausible explanation for the difference in size between eukaryotic and prokaryotic Okazaki fragments is that eukaryotic DNA replication occurs within the context of chromatin.

To investigate the relationship between lagging-strand processing and chromatin assembly, we have developed strategies to analyse Okazaki fragments purified from *S. cerevisiae*. We find that Okazaki fragments are heterogeneously sized with a repeat length corresponding to that of nucleosomes. In addition, we find that ligation junctions occur preferentially around known nucleosome midpoints, and that these relationships can be altered by interfering with chromatin assembly or Pol δ processivity. Our data suggest an integrated model whereby nascent chromatin structure facilitates Okazaki fragment processing, which may provide a means to monitor nucleosome assembly behind the replication fork.

Nucleosome-sized Okazaki fragments

To enrich for ligation-competent Okazaki fragments, we constructed strains of *S. cerevisiae* in which DNA ligase I¹¹ (*CDC9*) expression is driven from a doxycycline-repressible promoter and the encoded protein tagged with a temperature-sensitive degron¹⁶. The inability to ligate Okazaki fragments should result in the accumulation of nicked DNA that, after purification, can be radiolabelled using DNA polymerase and α -³²P dCTP. Ligase inactivation in an asynchronous culture resulted in the accumulation of short heterogeneous DNA species whose abundance increased with duration of ligase inhibition (Fig. 1a). Notably, these fragments displayed an underlying periodicity reminiscent of the nucleosome repeat. Indeed, comparison of the nicked DNA to a nucleosome ladder indicated that a significant fraction of fragments seem to be sized according to the 165-bp nucleosomal repeat length in *S. cerevisiae* (Fig. 1b).

To clarify that we were observing Okazaki fragments, we confirmed that they did not appear in a G1-arrested culture until S-phase release (Fig. 1c). Additionally, we treated purified DNA with recombinant DNA ligase; this repaired the nicks and severely diminished our ability to label fragments (Fig. 1d), showing that our assay predominantly detects nicked DNA rather than species containing single-stranded gaps or flaps resulting from incomplete Okazaki fragment processing. Unrelated control experiments indicated that a 125-nucleotide species previously reported to be Okazaki fragments^{14,17} is 5S ribosomal RNA (Supplementary Fig. 2). Thus, we demonstrate that the mono-, di- and

¹Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA.

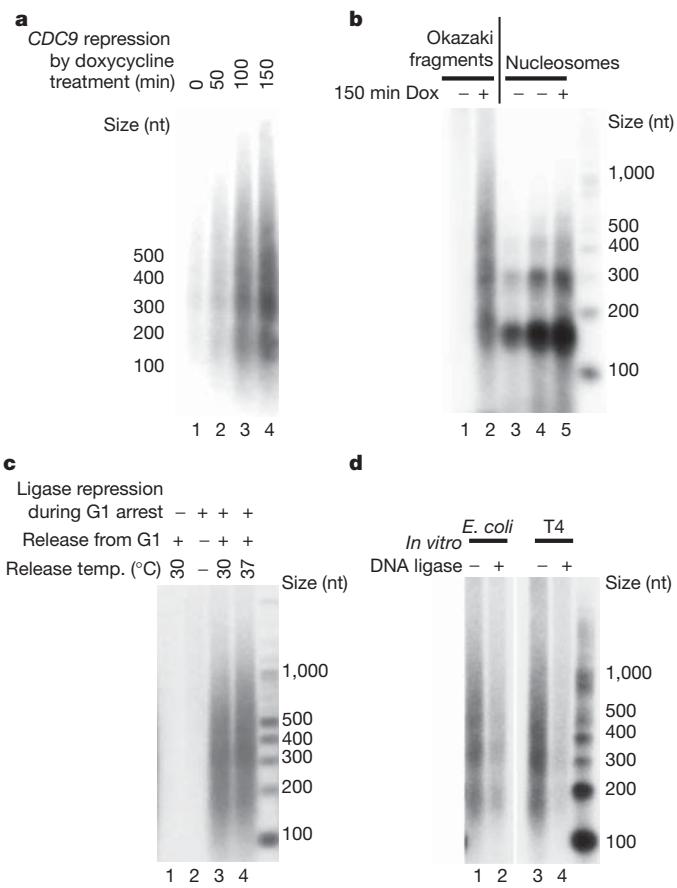


Figure 1 | DNA ligase I depletion in *S. cerevisiae* leads to the accumulation of Okazaki fragments sized similarly to the nucleosome repeat.

a, Transcriptional repression of DNA ligase I (*CDC9*) results in the accumulation of nicked DNA. Cells carrying a doxycycline-repressible allele of the *CDC9* gene were treated with doxycycline (Dox) for the indicated time. Purified genomic DNA was labelled using exonuclease-deficient Klenow fragment and α -³²P dCTP and separated in a denaturing agarose gel. nt, nucleotides. **b**, The size of labelled Okazaki fragments mirrors the nucleosome repeat. Okazaki fragments (lane 2) were labelled as in Fig. 1a; nucleosomes were prepared from wild-type (lane 3) or repressible *CDC9* strains (lanes 4 and 5) by *in vivo* micrococcal nuclease (MNase) digestion. The chromatin digestion patterns in lanes 4 and 5 indicate that *CDC9* repression does not alter global chromatin structure. **c**, Okazaki fragments accumulate during S phase. Cells were arrested in G1 using α -factor, during which time *CDC9* transcription was inhibited by the addition of doxycycline, and degradation of the protein stimulated by activation of the degron system using galactose and 37 °C. Okazaki fragments appear upon release of the culture into S phase (lanes 3 and 4). **d**, Okazaki fragments are bordered by ligatable nicks. Purified DNA was treated (lanes 2 and 4) or mock-treated (lanes 1 and 3) with the indicated ligase, and then labelled as in Fig. 1a. The inability to label fragments after ligase treatment confirms that labelled Okazaki fragments are flanked by ligatable nicks.

trinucleosome-sized fragments observed on ligase inhibition are generated in S phase and bordered by ligatable nicks—the essential properties of ligation-competent Okazaki fragments.

Global distribution of Okazaki fragments

To explore further the relationship between Okazaki fragments and nucleosomes, we developed a deep sequencing approach to map the strand, position and abundance of fragments across the *S. cerevisiae* genome. To allow yeast to complete S phase in the presence of nicked DNA we inactivated the DNA damage checkpoint by deleting the *RAD9* gene. Additionally, to deplete ligase activity further we deleted the second DNA ligase (*DNL4*). *DNL4* or *RAD9* deletion does not affect Okazaki fragment size (Supplementary Fig. 3). Okazaki fragments were harvested from an asynchronous culture after a ~2.5 h

ligase inactivation. Small single-stranded fragments were purified by anion exchange chromatography in alkaline conditions and sequencing primers were ligated directly to each end¹⁸ (Supplementary Fig. 4). Importantly, this method preserves strand identity.

After paired-end deep sequencing, we aligned the data to the *S. cerevisiae* genome and found that use of a single asynchronous culture was sufficient to attain complete coverage (a representative chromosome is shown in Fig. 2a, and a comparison of replicates in Supplementary Fig. 5). Sequencing reads aligning to the Watson or Crick strands showed a complementary distribution with strand bias being particularly pronounced close to replication origins¹⁹; such asymmetry is expected, given that replication forks proceed bidirectionally from origins (Fig. 2b). The strong strand bias observed around experimentally validated replication origins mapped at high resolution (Fig. 2c)²⁰ demonstrates preferential sequencing of nascent lagging strands, unequivocally confirming that the DNA species enriched after ligase repression are Okazaki fragments. Detailed analysis of global Okazaki fragment distributions can provide mechanistic insight into replication-fork dynamics (S. McGuffee, D.J.S. and I.W., manuscript in preparation).

Okazaki fragments terminate within nucleosomes

The periodic size of Okazaki fragments observed in Fig. 1 led us to investigate how the ends of Okazaki fragments relate to nucleosome positions found within the yeast genome. We aligned Okazaki fragment termini to a reference list of all consensus nucleosome midpoints (dyads) in *S. cerevisiae*²¹. Inherent bias towards smaller fragments during purification, library amplification and sequencing result in size distributions that differ substantially between starting material and sequenced fragments (Supplementary Fig. 6); therefore, we randomly selected fragments from our sequencing data to approximate the mononucleosome-sized fragments we detected by electrophoresis (Supplementary Fig. 6). Aligning both the 5' and 3' ends of mature (Fig. 3a) mononucleosome-sized fragments against the dyads of

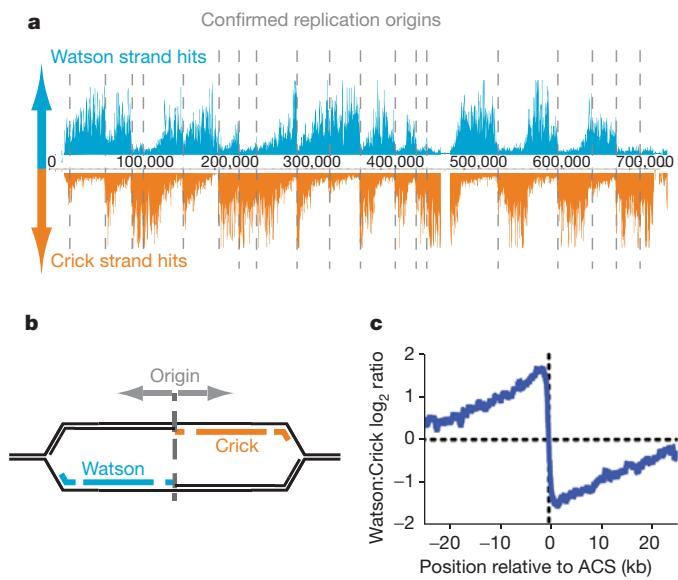


Figure 2 | Sequenced Okazaki fragments show a pronounced bias towards the lagging strand. **a**, Distribution of paired-end sequencing hits mapping to either the Watson (blue, above the axis) or Crick (orange, below) strands across *S. cerevisiae* chromosome 10. The locations of replication origins¹⁹ are indicated as grey dashed lines. Data are unsmoothed. **b**, The anticipated distribution of Okazaki fragments surrounding an efficient origin. **c**, Sequenced Okazaki fragments are strongly enriched in regions predicted to be replicated as the lagging strand. Log₂ ratio of Watson strand: Crick strand fragments across a 50-kb window around ARS consensus sequences (ACS) confirmed previously²⁰ to correspond to active origins. Data are smoothed to 200 bp.

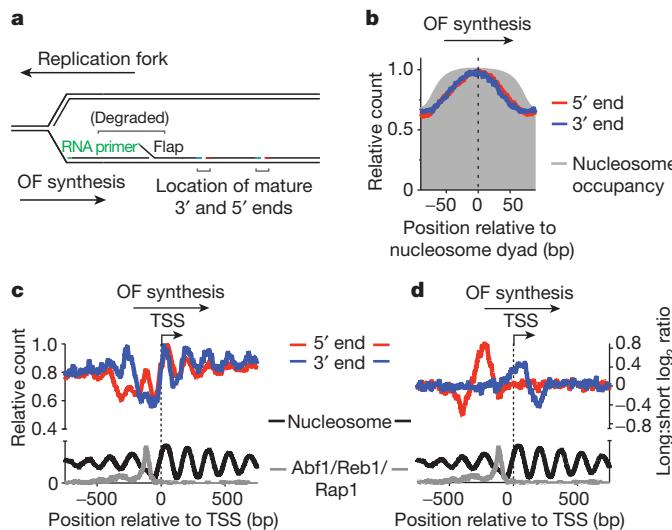


Figure 3 | Okazaki fragment termini are preferentially located at nucleosome dyads. **a**, Schematic indicating the mature Okazaki fragment termini sequenced in this study. **b**, Okazaki fragment termini are enriched around nucleosome dyads. The distribution of termini from mononucleosome-sized Okazaki fragments around the top 50% consensus *S. cerevisiae* nucleosome dyad locations²¹ is shown. Unless otherwise indicated, all analyses use unsmoothed data normalized to the maximum signal in the analysed range, and area aligned such that Okazaki fragment (OF) synthesis by Pol δ proceeds from left to right. **c**, The distribution of Okazaki fragment termini correlates with nucleosome occupancy²⁸ around TSSs. Data are aligned to TSSs such that the direction of transcription is from left to right: only Okazaki fragments synthesized in the same direction as transcription are analysed here; equivalent analyses for fragments synthesized in the opposite direction can be found in Supplementary Fig. 7. Data are smoothed to 5 bp. **d**, Okazaki fragments at nucleosome-depleted regions are disproportionately likely to be long. Equal numbers of reads were selected from the top (long), and the bottom three (short) length quartiles of sequenced fragments. The \log_2 ratio of long to short fragments is plotted. Data are smoothed to 5 bp.

high-confidence nucleosomes indicated that the highest density of ends occurs at the dyad (Fig. 3b). Alignment of unfiltered data and dinucleosome-sized fragments is comparable to the mononucleosome size-selected subset (Supplementary Fig. 6). Therefore, rather than being overrepresented in the accessible internucleosomal linker DNA, we find that Okazaki fragment termination is most likely to occur at locations corresponding to nucleosome dyads.

DNA-bound proteins dissociate Pol δ

To confirm that the observed correspondence between nucleosome occupancy and Okazaki fragment end density was not biased by a small subset of nucleosomes, we analysed transcription start sites (TSSs), at which nucleosome positioning follows a well-defined pattern²². Alignment of Okazaki fragment termini around \sim 4,500 TSSs indicated a marked correlation between nucleosome occupancy and the density of Okazaki fragment ends (Fig. 3c and Supplementary Fig. 7) across a broad region surrounding TSSs. Precise nucleosome positioning around TSSs is determined by chromatin remodelling^{22,23}, allowing us to infer that the correlation refers to remodelled, rather than intrinsically preferred, nucleosome locations.

If the ends of Okazaki fragments are dictated by nucleosomes then nicks should always be in register with positioned nucleosomes, regardless of internucleosomal spacing. Thus, nucleosomes separated by long linker DNA should be associated with longer Okazaki fragments. We took advantage of the differential nucleosome spacing surrounding the nucleosome-depleted region (NDR) at TSSs, comparing the distribution of equal numbers of randomly selected long and short fragments (defined, respectively, as the top quartile of sequenced fragments by length, and the bottom three quartiles).

When the ratio of long to short fragments is plotted across this region (Fig. 3d and Supplementary Fig. 7), we find that fragments whose ends map within nucleosomes bordering the NDR are disproportionately long. Thus, Okazaki fragment length reflects nucleosome spacing.

As well as correlating with nucleosome occupancy, Okazaki fragment termini were enriched around known transcription factor binding sites (Fig. 3c)^{24,25}. This suggested that, in addition to nucleosomes, sequence-specific DNA binding factors might directly influence the ends of Okazaki fragments. Such correlations would be expected if, following replication-fork passage, some transcription factors rapidly bound to replicated DNA and impeded Pol δ during strand-displacement synthesis, causing the polymerase to dissociate and leave a nick close to the site of collision. Furthermore, this model would also provide an explanation for the observed link between nucleosome occupancy and Okazaki fragment termination: nucleosomes contain numerous weak protein-DNA interactions, whose cumulative strength peaks around the dyad²⁶. Polymerases can invade nucleosomes, but experience increasing resistance as they approach the dyad^{27,28}. Therefore, if nucleosomes are rapidly re-established on the lagging strand behind the replication fork, and Pol δ invades them during strand-displacement synthesis, the likelihood of polymerase stalling and/or dissociation will increase with proximity to the dyad (see model in Supplementary Fig. 1), producing Okazaki fragments sized according to the nucleosome repeat. Other models that invoke biased Okazaki fragment initiation due to replisome pausing on encountering nucleosomes would produce distinct correlations around NDRs (Supplementary Fig. 8).

If polymerization by Pol δ is inhibited by sequence-specific DNA-bound transcription factors, Okazaki fragment ends should accumulate on the side of the factor that is first encountered by Pol δ . Aligning all transcription factor binding sites²⁵ indicated a significant enrichment of precisely juxtaposed Okazaki fragment 5' and 3' termini around the replication-fork-proximal side of the binding site (Fig. 4a). However, when the Abf1, Reb1 and Rap1 binding sites were considered separately we found that, among transcription factors, these proteins most strongly biased the positioning of Okazaki fragment ends (Fig. 4b, c and Supplementary Fig. 9). Abf1, Reb1 and Rap1 have essential roles in chromatin organization^{24,29}, and may therefore bind more rapidly and tightly to newly replicated DNA than the majority of transcription factors. We observed a strand-dependent enrichment of Okazaki fragment termini around the 3' ends of tRNA genes, as well as transcription factor TFIIIB binding sites (Supplementary Fig. 10). Thus, it seems that diverse DNA binding proteins can influence the processing of Okazaki fragments and probably facilitate the dissociation of Pol δ , with tightly bound protein complexes stimulating termination more precisely than those, such as nascent histones, that have more diffuse binding regions.

Chromatin assembly dictates nick location

Two testable predictions arise from the model that Pol δ dissociates via interaction with nascent nucleosomes: (1) disrupting nucleosome

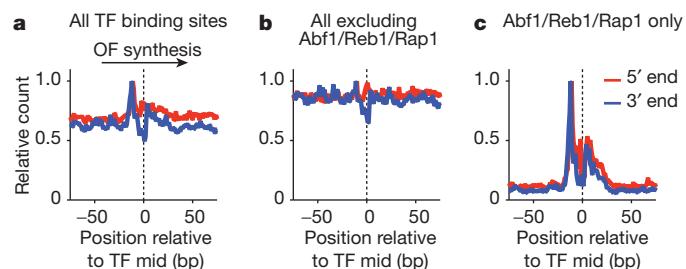


Figure 4 | Transcription factors with roles in nucleosome positioning stimulate dissociation of Pol δ . **a**, Okazaki fragment termini are enriched on the replication-fork-proximal side of known transcription factor (TF) binding sites²⁵. Mid, midpoint of TF binding site. **b, c**, The enrichment of Okazaki fragment termini around transcription factor binding sites can be attributed almost entirely to three factors—Abf1, Reb1 and Rap1—known to have roles in nucleosome positioning.

assembly will impair nucleosome-mediated Okazaki fragment termination and should, therefore, alter the size distribution of the fragments; and (2) reducing Pol δ processivity will result in a shift of Okazaki fragment termination sites away from the dyad, towards the replication-fork-proximal side of nucleosomes.

The deposition of (H3H4)₂ tetramers on nascent DNA—the first stage in nucleosome assembly—is mediated in part by CAF-1, a multi-subunit histone chaperone complex associated with the replisome via PCNA³⁰. We deleted each subunit of CAF-1 individually, and compared the length of Okazaki fragments in the deletion strains to those with wild-type CAF-1 (Fig. 5a). Loss of any CAF-1 subunit abrogated the periodic sizing of Okazaki fragments, significantly increasing their average length. These data are consistent with a global delay in (H3H4)₂ deposition following replication, with the resulting paucity of tetrasomes/nucleosomes on the lagging strand leading to increased strand displacement by Pol δ and thus to longer Okazaki fragments.

In *S. cerevisiae*, Pol δ is a heterotrimer whose Pol32 subunit increases processivity and PCNA binding affinity³¹; Pol δ lacking Pol32 shows decreased strand-displacement synthesis *in vitro* and probably *in vivo*³². Labelled Okazaki fragments from a *pol32* Δ strain were more heterogeneous but were still generally sized according to the nucleosome repeat (Fig. 5b). When the positions of the 5' and 3' termini of mononucleosome-sized fragments (165 \pm 15 nucleotides, as per Fig. 3b) from the *pol32* Δ strain were aligned against nucleosome midpoints, a clear shift was observed towards the first edge of the nucleosome encountered by the polymerase (Fig. 5c and Supplementary Fig. 11). Therefore, when Pol δ processivity is perturbed, the polymerase dissociates more rapidly due to an inability to invade nascent nucleosomes assembled on Okazaki fragments. The new maximum end density occurs \sim 35–40 nucleotides from the dyad, consistent with the extent of DNA protected by (H3H4)₂ tetramers³³: nucleosome assembly occurs step-wise, with assembly of (H3H4)₂ tetramers preceding recruitment of H2A/H2B³⁴; it is therefore possible that the histone-DNA species encountered by Pol δ is a tetrasome rather than a nucleosome, although our data do not allow us to distinguish between these possibilities. Importantly, the distribution of Okazaki fragment termini around transcription factor binding sites was identical between wild-type and *pol32* Δ strains (Supplementary Fig. 9), allowing us to conclude that the shift towards the edge of nucleosomes does not simply

represent a constant decrease of \sim 35–40 bp in the extent of strand displacement.

Discussion

Our data suggest a mechanism by which Pol δ may reliably remove DNA synthesized by Pol α while avoiding excessive strand-displacement synthesis. Repeated cycles of extension and cleavage can occur during the synthesis of a single Okazaki fragment. After removal of the RNA primer, all DNA flap structures generated by Pol δ are biochemically indistinguishable from one another, and an external mechanism is therefore required to measure the extent of strand displacement already carried out by Pol δ on each fragment. The removal of Pol δ by newly deposited histones represents a simple way to constrain Pol δ extension, and might allow nucleosome assembly to stimulate replication-fork progression directly. Although our data imply that nucleosomes strongly impede Pol δ , we note that both Fen1 and DNA ligase I can act efficiently on nucleosomal substrates^{35,36}. We speculate that carrying out Okazaki fragment processing in the context of nucleosomes rather than DNA allows the ligation reaction to outcompete further strand displacement by Pol δ .

CAF-1 is not essential in *S. cerevisiae*, suggesting that other histone chaperones such as Asf1, Rtt106 and HIR³⁷ allow sufficient replication-coupled or post-replicative nucleosome assembly for viability. Nevertheless, our results are consistent with a temporal delay in nucleosome assembly in CAF-1 mutants, which may give rise to previously reported silencing defects³⁸. The importance of replication-coupled chromatin assembly in metazoa is illustrated by the severe phenotypes of CAF-1 or ASF1 disruption in human cell lines: ASF1 depletion leads to replication-fork stalling³⁹, and CAF-1 depletion also precludes progression through S phase⁴⁰.

Our studies demonstrate that the location of Okazaki fragment termini can be determined by interactions between the lagging-strand polymerase and nascent nucleosomes. This observation provides the first direct mechanistic evidence for the coupling of DNA replication to chromatin assembly on the newly replicated daughter genomes. Such coupling is fundamentally important given the part played by chromatin in the regulation of gene expression, as well as the potential for the epigenetic inheritance of precisely located modified nucleosomes. In addition, transcription factors known to have roles in the establishment of chromatin structure are apparently able to re-bind to DNA immediately after replication-fork passage. Unlike histones, which are present in sufficient number to be distributed to both daughter genomes, DNA-bound transcription factors are present at only one local copy per two daughters. Regulation of transcription factor re-binding could thus have a key role in asymmetric epigenetic inheritance.

METHODS SUMMARY

Genomic DNA was prepared from spheroblasts essentially as described for medium resolution DSB mapping⁴¹, labelled using exo-Klenow fragment (NEB) and α -³²P dCTP and separated in 1.3% denaturing agarose gels.

Okazaki fragments were purified by binding heat-denatured genomic DNA to Source 15Q (GE Healthcare) at pH 12, 300 mM NaCl, and eluting in 50 mM steps to 1 M. Primers for paired-end Illumina TruSeq sequencing were directly ligated to each end. Following second-strand synthesis and size selection by gel purification, libraries were amplified by PCR, purified from two sequential 2.5% native agarose gels, and sequenced directly. Full details are in Supplementary Fig. 4 and Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 November 2011; accepted 24 January 2012.

Published online 14 March 2012.

- Corpet, A. & Almouzni, G. Making copies of chromatin: the challenge of nucleosomal organization and epigenetic information. *Trends Cell Biol.* **19**, 29–41 (2009).

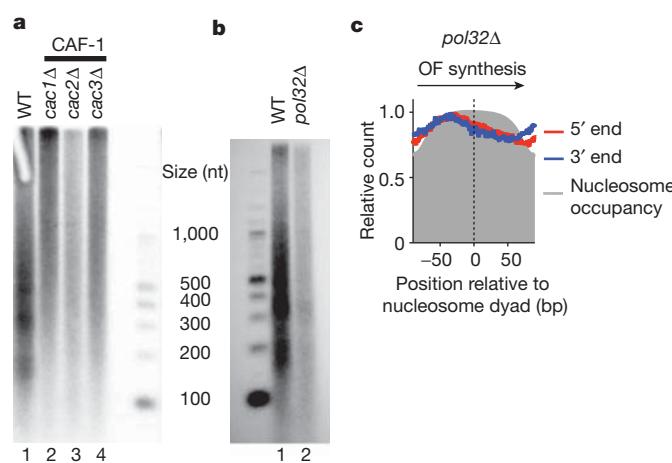


Figure 5 | Impaired chromatin assembly and Pol δ processivity affect the size of Okazaki fragments and the location of their termini, respectively. **a**, Deletion of any component of the CAF-1 complex (lanes 2–4) results in Okazaki fragments that are no longer sized according to nucleosomes. WT, wild type. **b**, Deletion of the *pol32* subunit of Pol δ (lane 2) does not abrogate the chromatin-like size distribution of Okazaki fragments. **c**, In the absence of Pol32, the termini of mononucleosome-sized Okazaki fragments are shifted towards the replication-fork-proximal edge of the nucleosome: peak density occurs at a location consistent with the predicted edges of (H3H4)₂ tetramers. See also Supplementary Fig. 11.

2. Sogo, J. M., Stahl, H., Koller, T. & Knippers, R. Structure of replicating simian virus 40 minichromosomes. The replication fork, core histone segregation and terminal structures. *J. Mol. Biol.* **189**, 189–204 (1986).
3. Burgers, P. M. Polymerase dynamics at the eukaryotic DNA replication fork. *J. Biol. Chem.* **284**, 4041–4045 (2009).
4. Kaufmann, G. & Falk, H. H. An oligoribonucleotide polymerase from SV40-infected cells with properties of a primase. *Nucleic Acids Res.* **10**, 2309–2321 (1982).
5. Nethanel, T. & Kaufmann, G. Two DNA polymerases may be required for synthesis of the lagging DNA strand of simian virus 40. *J. Virol.* **64**, 5912–5918 (1990).
6. Waga, S. & Stillman, B. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature* **369**, 207–212 (1994).
7. Ayyagari, R., Gomes, X. V., Gordenin, D. A. & Burgers, P. M. Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 and DNA2. *J. Biol. Chem.* **278**, 1618–1625 (2003).
8. Bae, S. H., Bae, K. H., Kim, J. A. & Seo, Y. S. RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature* **412**, 456–461 (2001).
9. Kao, H. I., Veeraraghavan, J., Polaczek, P., Campbell, J. L. & Bambara, R. A. On the roles of *Saccharomyces cerevisiae* Dna2p and Flap endonuclease 1 in Okazaki fragment processing. *J. Biol. Chem.* **279**, 15014–15024 (2004).
10. Garg, P., Stith, C. M., Sabouri, N., Johansson, E. & Burgers, P. M. Idling by DNA polymerase δ maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev.* **18**, 2764–2773 (2004).
11. Johnston, L. H. & Nasmyth, K. A. *Saccharomyces cerevisiae* cell cycle mutant *cdc9* is defective in DNA ligase. *Nature* **274**, 891–893 (1978).
12. Pavlov, Y. I. *et al.* Evidence that errors made by DNA polymerase α are corrected by DNA polymerase δ . *Curr. Biol.* **16**, 202–207 (2006).
13. Anderson, S. & DePamphilis, M. L. Metabolism of Okazaki fragments during simian virus 40 DNA replication. *J. Biol. Chem.* **254**, 11495–11504 (1979).
14. Bielinsky, A. K. & Gerbi, S. A. Discrete start sites for DNA synthesis in the yeast *ARS1* origin. *Science* **279**, 95–98 (1998).
15. Blumenthal, A. B. & Clark, E. J. Discrete sizes of replication intermediates in *Drosophila* cells. *Cell* **12**, 183–189 (1977).
16. Dohmen, R. J. & Varshavsky, A. Heat-inducible degron and the making of conditional mutants. *Methods Enzymol.* **399**, 799–822 (2005).
17. Bielinsky, A. K. & Gerbi, S. A. Chromosomal *ARS1* has a single leading strand start site. *Mol. Cell* **3**, 477–486 (1999).
18. Ng, P. *et al.* Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nature Methods* **2**, 105–111 (2005).
19. Nieduszynski, C. A., Hiraga, S., Ak, P., Benham, C. J. & Donaldson, A. D. OriDB: a DNA replication origin database. *Nucleic Acids Res.* **35**, D40–D46 (2007).
20. Eaton, M. L., Galani, K., Kang, S., Bell, S. P. & MacAlpine, D. M. Conserved nucleosome positioning defines replication origins. *Genes Dev.* **24**, 748–753 (2010).
21. Jiang, C. & Pugh, B. F. A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome. *Genome Biol.* **10**, R109 (2009).
22. Whitehouse, I., Rando, O. J., Delrow, J. & Tsukiyama, T. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**, 1031–1035 (2007).
23. Hartley, P. D. & Madhani, H. D. Mechanisms that specify promoter nucleosome location and identity. *Cell* **137**, 445–458 (2009).
24. Badis, G. *et al.* A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* **32**, 878–887 (2008).
25. MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**, 113 (2006).
26. Hall, M. A. *et al.* High-resolution dynamic mapping of histone–DNA interactions in a nucleosome. *Nature Struct. Mol. Biol.* **16**, 124–129 (2009).
27. Bondarenko, V. A. *et al.* Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol. Cell* **24**, 469–479 (2006).
28. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
29. Bai, L., Ondracka, A. & Cross, F. R. Multiple sequence-specific factors generate the nucleosome-depleted region on *CLN2* promoter. *Mol. Cell* **42**, 465–476 (2011).
30. Shibahara, K. & Stillman, B. Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* **96**, 575–585 (1999).
31. Johansson, E., Garg, P. & Burgers, P. M. The Pol32 subunit of DNA polymerase δ contains separable domains for processive replication and proliferating cell nuclear antigen (PCNA) binding. *J. Biol. Chem.* **279**, 1907–1915 (2004).
32. Stith, C. M., Sterling, J., Resnick, M. A., Gordenin, D. A. & Burgers, P. M. Flexibility of eukaryotic Okazaki fragment maturation through regulated strand displacement synthesis. *J. Biol. Chem.* **283**, 34129–34140 (2008).
33. Dong, F. & van Holde, K. E. Nucleosome positioning is determined by the (H3–H4)2 tetramer. *Proc. Natl. Acad. Sci. USA* **88**, 10596–10600 (1991).
34. Smith, S. & Stillman, B. Stepwise assembly of chromatin during DNA replication *in vitro*. *EMBO J.* **10**, 971–980 (1991).
35. Chafin, D. R., Vitolo, J. M., Henricksen, L. A., Bambara, R. A. & Hayes, J. J. Human DNA ligase I efficiently seals nicks in nucleosomes. *EMBO J.* **19**, 5492–5501 (2000).
36. Huggins, C. F. *et al.* Flap endonuclease 1 efficiently cleaves base excision repair and DNA replication intermediates assembled into nucleosomes. *Mol. Cell* **10**, 1201–1211 (2002).
37. Ray-Gallet, D. *et al.* Dynamics of histone H3 deposition *in vivo* reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol. Cell* **44**, 928–941 (2011).
38. Zhang, Z., Shibahara, K. & Stillman, B. PCNA connects DNA replication to epigenetic inheritance in yeast. *Nature* **408**, 221–225 (2000).
39. Groth, A. *et al.* Regulation of replication fork progression through histone supply and demand. *Science* **318**, 1928–1931 (2007).
40. Hoek, M. & Stillman, B. Chromatin assembly factor 1 is essential and couples chromatin assembly to DNA replication *in vivo*. *Proc. Natl. Acad. Sci. USA* **100**, 12183–12188 (2003).
41. Murakami, H., Borde, V., Nicolas, A. & Keeney, S. Gel electrophoresis assays for analyzing DNA double-strand breaks in *Saccharomyces cerevisiae* at various spatial resolutions. *Methods Mol. Biol.* **557**, 117–142 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. McGuffee for assistance with data processing; S. Keeney, K. Marians, D. Remus, T. Tsukiyama, members of the Molecular Biology Program and Whitehouse laboratory for discussions and comments on the manuscript. This work was supported by a Louis V. Gerstner Jr Young Investigator Award and an Alfred Bressler Scholars Endowment Award to I.W. D.J.S. is an HHMI fellow of the Damon Runyon Cancer Research Foundation (DRG-#2046-10).

Author Contributions D.J.S. and I.W. designed experiments; D.J.S. performed experiments and analysed data; D.J.S. and I.W. interpreted results; the manuscript was drafted by D.J.S. and edited by D.J.S. and I.W.

Author Information Raw sequencing data and processed data are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33786>) under accession number 33786. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to I.W. (whitehoi@mskcc.org).

METHODS

DNA purification. Yeast strains carrying degron-tagged, doxycycline-repressible alleles of *CDC9* and a galactose-inducible *UBR1* allele (see Supplementary Table 1 for a list of strains) were grown at 30 °C in YEP supplemented with 2% raffinose. At optical density (OD) 0.4, galactose and doxycycline were added to final concentrations of 2% and 40 mg l⁻¹, respectively, and the culture shaken at 37 °C for 2.5 h. Fifty-millilitre cultures were used for labelling experiments, and 250-ml cultures for purification and library generation.

Genomic DNA was prepared from spheroblasts as described for medium resolution DSB mapping⁴¹. Following ligase repression, cells were collected by centrifugation, washed in SCE buffer (1 M sorbitol, 100 mM sodium citrate, 60 mM EDTA, pH 7.0) and spheroblasted for 3 min with 5 mg zymolyase 20T (USB) per 50-ml culture. Spheroblasts were washed with SCE, and resuspended in 480 µl lysis buffer (50 mM Tris-HCl, pH 8.0, 50 mM EDTA, 100 mM NaCl, 1.5% sarkosyl) containing 150 µg proteinase K (Fisher). Digestion was carried out for 2–16 h at 37 °C. After digestion, residual proteins and peptides were precipitated by adding 200 µl 5 M KOAc and spinning at 16,000g for 30 min at 4 °C. Nucleic acids were precipitated from the supernatant by addition of 500 µl isopropanol and centrifugation at 16,000g for 10 min. Pellets were washed twice with 500 µl 70% ethanol, resuspended in 200 µl STE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl) and digested with 25 µg RNase A (Sigma) and/or 10 U Riboshredder RNase blend (Epicentre) at 37 °C for 30 min. Genomic DNA was precipitated by addition of 20 µl NaOAc, pH 5.5 and 800 µl ethanol followed by centrifugation at 5,000g for 10 min at room temperature (25 °C). Pellets were washed with 70% ethanol and resuspended in 1 µl TE (10 mM Tris:Cl pH 7.5, 0.1 mM EDTA) per ml original culture volume. DNA was stored at 4 °C and never frozen.

DNA labelling. Two microlitres of DNA (corresponding to the genomic DNA content of 2 ml cultured cells) was used in 20 µl labelling reactions containing 5 U Klenow (exo-)polymerase (NEB) and α -dCTP (Perkin Elmer) at a final concentration of 33 nM. Free label was removed using Illustra microspin G-50 columns (GE healthcare). Labelled DNA was separated in 1.3% denaturing agarose gels (50 mM NaOH, 1 mM EDTA). After electrophoresis, the gel was neutralized and DNA transferred to an uncharged nitrocellulose membrane (Hybond-N; GE healthcare) via capillary transfer. Membranes were exposed to phosphor screens or film.

Okazaki fragment purification. Genomic DNA purified as described earlier was denatured by heating to 95 °C for 5 min, rapidly cooled on ice and brought to 300 mM NaCl, pH 12. Purification was carried out in batch using 400 µl Source 15Q (GE healthcare), binding at 300 mM NaCl, pH 12 and eluting in 50 mM steps to 1 M NaCl, pH 12. As determined by purification of fragments pre-labelled as above (see also Supplementary Fig. 4), fractions from 800–900 mM NaCl contained the majority of fragments of interest. DNA was ethanol precipitated and treated with 10 U Riboshredder RNase blend for 30 min at 37 °C to remove residual, undigested RNA; digestion products were removed using Illustra microspin G-50 columns to leave essentially pure Okazaki fragments.

Sequencing library generation. Adaptor primer pairs with single-stranded overhangs (shown schematically in Supplementary Fig. 4) were annealed by cooling from 95 °C and purified from 12% native polyacrylamide gels via standard methods. Sequences of the adaptor pairs are as follows. 5' top, ACACCTTTCCCTACACG ACGCTCTCCGATCT; 5' bottom, NNNNNNAGATCGGAAGAGCGTCGTGT AGGGAAAGAGTGT; 3' top, /Phos/AGATCGGAAGAGCGGTTCAAGCAGGAA TGCCGAG; 3' bottom, CTCGGCATTCTGCTGAACCGCTCTCCGATCTN NNNNN.

Up to 200 ng denatured purified Okazaki fragments were incubated at 16 °C overnight in a ligation reaction containing 1 µg of each primer pair and 1,000 U T4 DNA ligase (NEB). Unligated adaptors were removed using Illustra microspin S-300 columns (GE healthcare) and a second strand-synthesis reaction carried out at 72 °C using Taq polymerase (NEB). Products from ~200–1,000 bp were purified from 2.5% agarose gels run in TBE using Qiaquick kits (Qiagen). Purified libraries were amplified (16 cycles) using Illumina Truseq primers according to Illumina protocols, except that KOD hot start polymerase (Novagen) was used. Amplified libraries were purified from two sequential 2.5% agarose gels. Subsequent steps in the sequencing workflow were carried out according to standard procedures.

Nucleosome dyad and occupancy data. The top 50% of nucleosome dyads (by confidence score), taken from a meta-analysis²¹, were used in our analysis; nucleosome occupancy data (Fig. 3b) were from the YPD data set detailed previously⁴².

42. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).

Deciphering a neuronal circuit that mediates appetite

Qi Wu¹†, Michael S. Clark² & Richard D. Palmiter¹

Hypothalamic neurons that co-express agouti-related protein (AgRP), neuropeptide Y and γ -aminobutyric acid (GABA) are known to promote feeding and weight gain by integration of various nutritional, hormonal, and neuronal signals^{1,2}. Ablation of these neurons in mice leads to cessation of feeding that is accompanied by activation of *Fos* in most regions where they project^{3–6}. Previous experiments have indicated that the ensuing starvation is due to aberrant activation of the parabrachial nucleus (PBN) and it could be prevented by facilitating GABA_A receptor signalling in the PBN within a critical adaptation period⁵. We speculated that loss of GABA signalling from AgRP-expressing neurons (AgRP neurons) within the PBN results in unopposed excitation of the PBN, which in turn inhibits feeding. However, the source of the excitatory inputs to the PBN was unknown. Here we show that glutamatergic neurons in the nucleus tractus solitarius (NTS) and caudal serotonergic neurons control the excitability of PBN neurons and inhibit feeding. Blockade of serotonin (5-HT₃) receptor signalling in the NTS by either the chronic administration of ondansetron or the genetic inactivation of *Tph2* in caudal serotonergic neurons that project to the NTS protects against starvation when AgRP neurons are ablated. Likewise, genetic inactivation of glutamatergic signalling by the NTS onto N-methyl D-aspartate-type glutamate receptors in the PBN prevents starvation. We also show that suppressing glutamatergic output of the PBN reinstates normal appetite after AgRP neuron ablation, whereas it promotes weight gain without AgRP neuron ablation. Thus we identify the PBN as a hub that integrates signals from several brain regions to bidirectionally modulate feeding and body weight.

Administration of diphtheria toxin (DT) to *AgRP*^{DTR} mice, which express the human DT receptor selectively in AgRP neurons, ablates nearly all AgRP neurons in the arcuate nucleus of the hypothalamus; during the next 6 days the mice gradually cease eating, lose body weight, and die without intervention⁴. Chronic infusion of bretazenil, a partial agonist of GABA_A receptor, into the PBN during the DT treatment prevents starvation and allows an adaptive process to take place such that the mice eat and maintain their body weight⁵. Not only does ablation of AgRP neurons inhibit the initiation of meals, it also decreases the amount of liquid food that will be swallowed when it is delivered directly into the mouth⁷. Because the PBN responds to visceral malaise, such as food poisoning and LiCl treatment⁸, and also processes gustatory signals in paradigms such as the conditional taste aversion or preference^{9,10}, we predicted that ablation of AgRP neurons results in unopposed activation of the PBN, which may mimic a nausea signal and thereby inhibit feeding. To test this hypothesis, we infused ondansetron, an anti-nausea drug that antagonizes 5-HT₃ receptors¹¹, subcutaneously or directly into the fourth ventricle, starting 3 days before injection of *AgRP*^{DTR} mice with DT. Despite the fact that the drug is administered orally to people, only central delivery of ondansetron prevented fatal weight loss and allowed the mice to recover (Fig. 1a and Supplementary Fig. 1a). Consumption of low-fat chow pellets by ondansetron-treated mice fell and they lost roughly 10% of their body weight during the first week after DT treatment; however, they then gradually ate more and

regained body weight by 3 weeks after DT treatment (Fig. 1a and Supplementary Fig. 1a). The 5-HT₃ receptor is an excitatory ion channel that is expressed widely in the brain, especially in the cortex and dorsal brainstem¹². To examine more precisely where ondansetron acts to prevent starvation after AgRP neuron ablation, the drug was delivered bilaterally to either the PBN or the NTS (see Supplementary Fig. 2 for cannula placement). Delivery of ondansetron to the PBN did not rescue the starvation phenotype of DT-treated mice, whereas delivery to the NTS prevented starvation (Fig. 1b and Supplementary Fig. 1b). The results suggest that serotonin provides some of the excitatory drive that indirectly results in hyperactivity of the PBN after loss of inhibitory input from AgRP neurons. Neurons in the NTS are known to send excitatory, glutamatergic inputs to the PBN^{13,14}. We therefore predicted that serotonin action on 5-HT₃ receptors in the NTS promotes hyperexcitation of the PBN, which can be measured as local *Fos* gene activation⁶. Consistent with this hypothesis, *Fos* induction in the PBN was significantly ameliorated by the administration of ondansetron in the NTS (Supplementary Fig. 3). We conclude that inhibition of 5-HT₃-mediated excitatory currents in the vicinity of the NTS prevents starvation after the ablation of AgRP neurons and promotes an adaptation that allows feeding to be maintained in the absence of AgRP neurons.

Tryptophan hydroxylase 2 (*Tph2*) catalyses the first and rate-limiting step in serotonin biosynthesis in the central nervous system¹⁵. To examine the role of serotonin more directly, conditional *Tph2*^{lox/lox} mice carrying the *AgRP*^{DTR/+} allele were generated and then injected with *CAV2-Cre*, a virus that is retrogradely transported from the site of injection to the cell bodies where it can inactivate the *Tph2* gene only in those neurons that project to the injection site. *CAV2-Cre* was injected bilaterally into the NTS of *Tph2*^{lox/lox}; *AgRP*^{DTR/+} mice, and 8 days later they were treated with DT to ablate the AgRP neurons. This viral treatment prevented the starvation that normally occurs after ablation of AgRP neurons. Feeding and body weight decreased slightly after DT treatment of the virally transduced mice, but some of the mice restored normal food intake and regained body weight (Fig. 1c and Supplementary Fig. 1c). Various raphe nuclei from the virally rescued *Tph2*^{lox/lox} mice were examined for serotonergic neurons that lacked *Tph2* but retained L-aromatic amino acid decarboxylase (AADC), another marker of serotonergic cells. Many serotonergic cell bodies in the raphe obscurus (ROb) and raphe magnus (RMg) were found that significantly lacked *Tph2* staining (Fig. 2a–l). Quantification of the results revealed that viral treatment decreased *Tph2* signalling in ROb and RMg serotonergic neurons by 60–80% (Fig. 2m), whereas serotonergic neurons in the dorsal raphe (DR) were unaffected (Fig. 2m, and data not shown). Viral injection into the NTS of *Tph2*^{lox/lox} mice decreased serotonin levels in the NTS by about 75% (wild-type, 11.65 ± 0.92 ng per mg of protein; virus-injected, 2.68 ± 0.57 ng per mg of protein). Our observations are consistent with known projections of the caudal raphe neurons to brainstem structures and projections of DR neurons to forebrain regions¹⁶.

Because many of the neurons in the NTS are known to send glutamatergic projections to the PBN^{13,14}, we predicted that serotonergic

¹Howard Hughes Medical Institute and Department of Biochemistry, University of Washington School of Medicine, Seattle, Washington 98195, USA. ²Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington 98195, USA. [†]Present address: Eagles Diabetes Research Center and Department of Pharmacology, Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA.

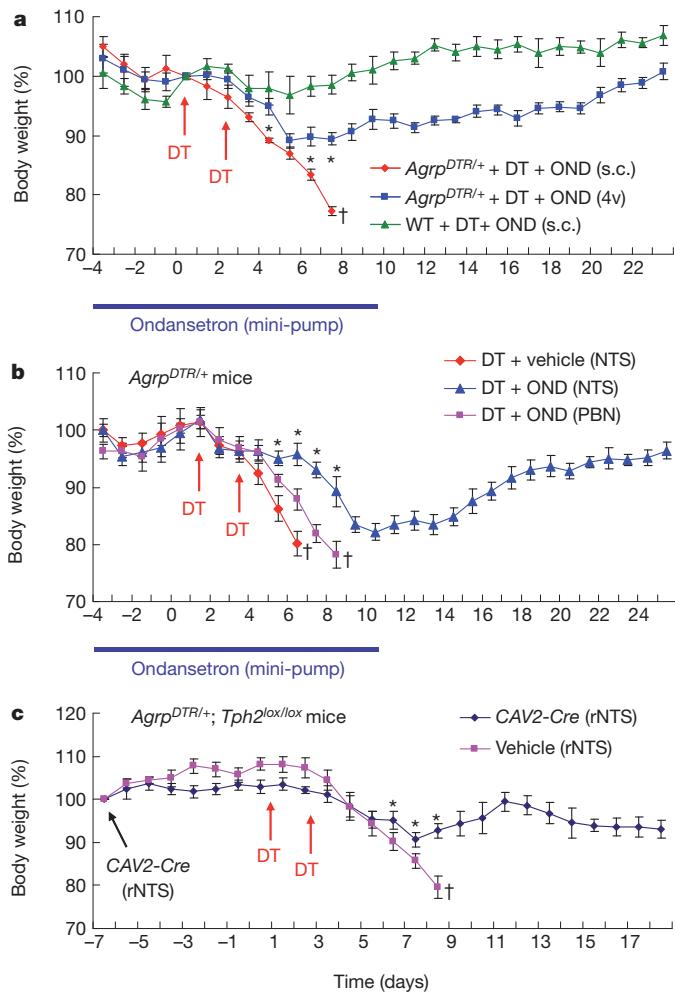


Figure 1 | Chronic administration of ondansetron into the NTS, or genetic inactivation of serotonergic input to the NTS prevents starvation in AgRP neuron-ablated mice. **a**, Body weight of *Agrp*^{DTR/+} mice and wild-type (WT) mice after either subcutaneous (s.c.; 1 mg kg⁻¹ d⁻¹, $n = 6$) or fourth-ventricle (4v; 0.1 mg kg⁻¹ d⁻¹, $n = 12$) infusion of ondansetron, a 5-HT₃ receptor antagonist, through osmotic minipumps. DT was injected intramuscularly twice as indicated by arrows. Minipumps were removed on day 11 after the pump content was depleted. Eight of 12 mice survived the treatment of DT and infusion of ondansetron into 4v (blue line). Asterisk, $P < 0.01$ between *Agrp*^{DTR/+} mice treated with ondansetron s.c. ($n = 6$ non-survivors) and 4v ($n = 8$ survivors). **b**, Body weight of DT-treated *Agrp*^{DTR/+} mice after chronic infusion of ondansetron or vehicle into either the NTS ($n = 14$) or the PBN ($n = 8$). DT was twice injected intramuscularly as indicated by arrows. Six of 14 mice survived the treatment of DT and infusion of ondansetron into the NTS (blue line). Asterisk, $P < 0.01$ between *Agrp*^{DTR/+} mice infused by ondansetron into the NTS ($n = 6$ survivors) and the PBN ($n = 8$ non-survivors). **c**, Body weight of DT-treated, *Agrp*^{DTR/+}; *Tph2*^{lox/lox} mice after bilateral injection of CAV2-Cre virus or vehicle into the rostral part of the NTS (rNTS). Seven of 12 mice survived the treatment with DT and viral infection (blue line). Asterisk, $P < 0.05$ between mice injected with CAV2-Cre ($n = 7$ survivors) and vehicle into the NTS ($n = 8$ non-survivors). Daggers indicate mice that were removed from the experiment when they either lost 20% of their body weight or seemed moribund. Results are shown as means \pm s.e.m.

activation of the NTS might lead to glutamatergic activation of the PBN, which could be responsible for starvation after AgRP neuron ablation. To test this hypothesis, we inactivated the vesicular glutamate transporter 2 (Vglut2, encoded by the *Slc17a6* gene) within the NTS by injecting *AAV1-CreGFP* virus bilaterally into the NTS of *Slc17a6*^{lox/lox}; *Agrp*^{DTR/+} mice 8 days before the initiation of DT treatment (see Supplementary Fig. 4 for the placement of virus injections). Mice that were injected bilaterally and treated with DT maintained body weight and feeding, whereas mice that were unilaterally injected starved, as did mice with vehicle

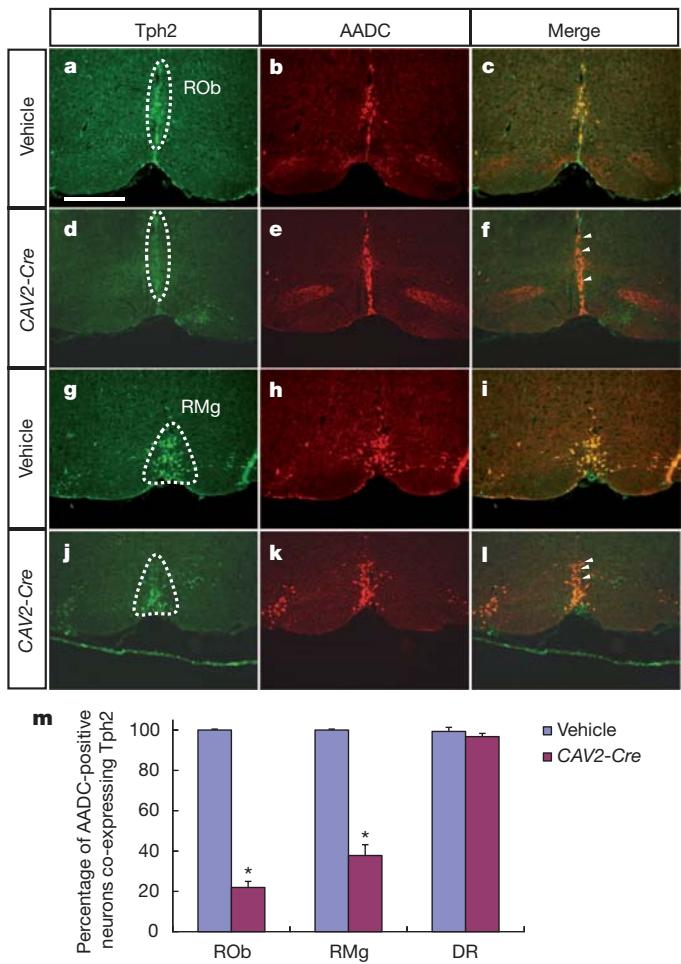


Figure 2 | Serotonergic projections from the ROb and RMg to the NTS mediate starvation after ablation of AgRP neurons. **a–f**, Representative immunohistochemistry pictures of Tph2 and AADC, markers of serotonergic neurons, in *Agrp*^{DTR/+}; *Tph2*^{lox/lox} mice after bilateral injection of either vehicle (**a–c**) or a retrograding CAV2-Cre virus (**d–f**) into the rostral NTS. Arrowheads indicate the serotonin neurons within the ROb (B2 group) in which the expression of *Tph2* was gone after viral infection. **g–l**, Representative immunostaining pictures of Tph2 and AADC at the RMg (B3 group) from the mice described above. Arrowheads indicate the serotonin neurons within the RMg in which the expression of *Tph2* was gone after viral infection. **m**, Quantified immunohistochemistry results of AADC-expressing neurons that co-localized with Tph2-expressing neurons in the ROb, RMg and DR of the mice described in **a–f** (and data not shown). Asterisk, $P < 0.01$, analysis of variance (ANOVA); $n = 6$ mice per group. Results are shown as means and s.e.m. Scale bar in **a** (for **a–l**), 400 μ m.

injection (Fig. 3a and Supplementary Fig. 5). To further establish that glutamatergic activation of the PBN inhibits feeding in this model, we used a viral/genetic approach to reduce *N*-methyl-D-aspartate (NMDA) receptors in the PBN and thereby dampen the excitability by glutamate. *Grin1*^{lox/lox}; *Agrp*^{DTR/+} mice, which carry two conditional alleles of the gene encoding the essential NR1 subunit of NMDA receptors and the *Agrp*^{DTR} allele, were injected bilaterally with *AAV1-CreGFP* in the PBN 9 days before ablation of the AgRP neurons by DT (see Supplementary Fig. 6 for the placement of virus injections). For the bilaterally injected mice, body weight declined slightly during the first 8 days after DT injection, but then recovered along with a rebound of food intake, whereas the vehicle-injected and unilaterally injected mice stopped eating and did not recover (Fig. 3b and Supplementary Fig. 7). These experiments show that either decreasing glutamatergic signalling by neurons within the NTS or decreasing the number of NMDA receptors in the PBN protects against the starvation caused by the ablation of AgRP neurons. Most neurons within the PBN are glutamatergic. We therefore predicted that

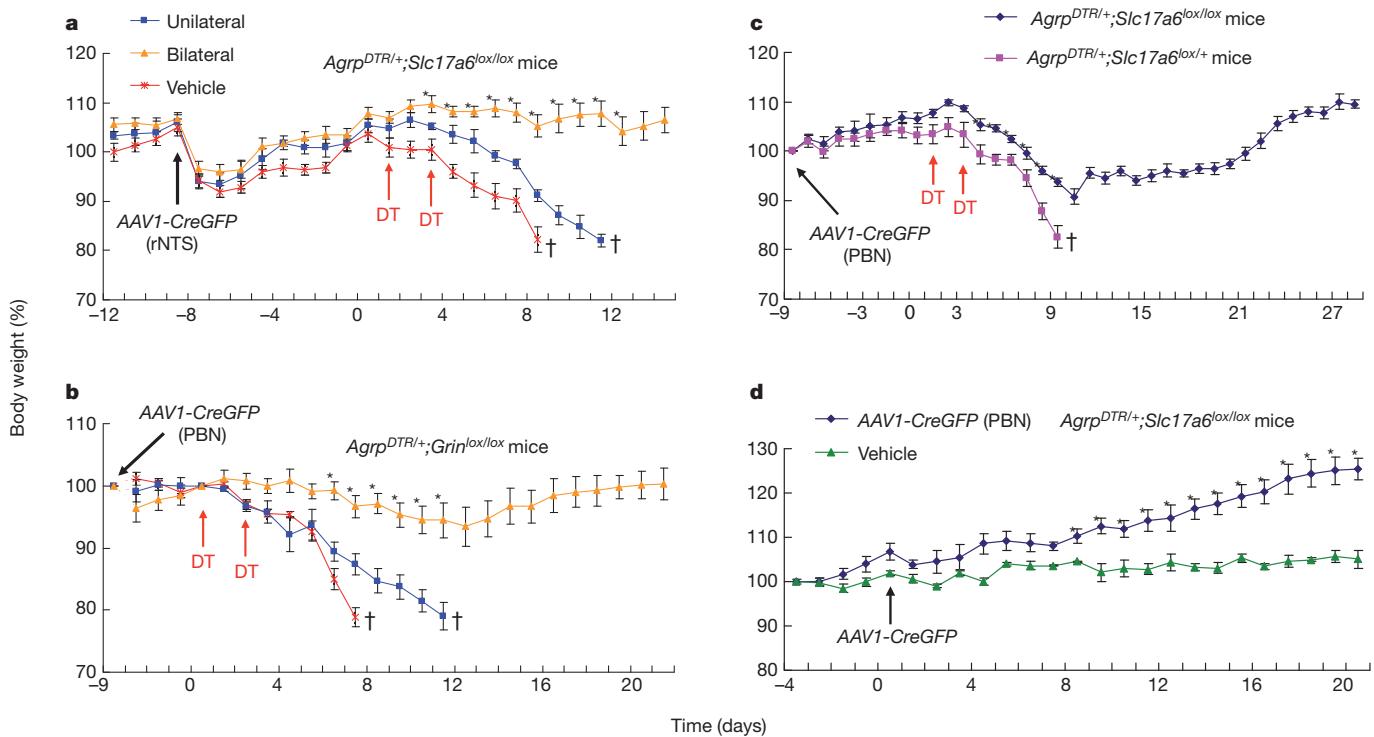


Figure 3 | Viral-mediated disruption of glutamatergic circuitry between the NTS and PBN, or glutamatergic output of the PBN, rescues feeding after ablation of AgRP neurons. **a**, Percentage of initial body weight of DT-treated $Agrp^{DTR/+}; Slc17a6^{lox/lox}$ mice after either bilateral ($n = 19$) or unilateral ($n = 9$) injection of AAV1-CreGFP virus or vehicle into the rostral NTS (rNTS; $n = 9$). AAV1-CreGFP virus decreases glutamatergic signalling from the rNTS to downstream targets, including the PBN. Out of 19 mice, 14 survived DT treatment and viral injections into the NTS (orange line). Asterisk, $P < 0.01$ between bilateral virus injection ($n = 14$ survivors) and unilateral virus injection ($n = 9$ non-survivors) or vehicle injection ($n = 9$ non-survivors). **b**, Percentage of initial body weight of DT-treated $Agrp^{DTR/+}; Grin1^{lox/lox}$ mice after either bilateral injection ($n = 18$) or unilateral injection of AAV1-CreGFP virus ($n = 8$) or vehicle ($n = 8$) into the lateral PBN. AAV1-CreGFP virus attenuates NMDA receptor signalling in the PBN, which receives dense glutamatergic projections from the NTS. Out of 18 mice, 11 survived DT treatment and viral injections into the PBN (orange line). Asterisk, $P < 0.01$

between bilateral virus injection ($n = 11$ survivors) and unilateral virus injection ($n = 8$ non-survivors) or vehicle injection ($n = 8$ non-survivors). **c**, Percentage of initial body weight of DT-treated $Agrp^{DTR/+}; Slc17a6^{lox/lox}$ mice ($n = 21$) or DT-treated $Agrp^{DTR/+}; Slc17a6^{lox/+}$ control mice ($n = 10$) after bilateral injection of AAV1-CreGFP virus into the lateral PBN. AAV1-CreGFP virus abolishes glutamatergic signalling from the PBN to various forebrain targets. Out of 21 mice, 12 survived DT treatment and viral injections into the PBN (blue lines). Asterisk, $P < 0.01$, ANOVA between 12 survivors of Vglut2-deficient group and 10 non-survivor control group. **d**, Percentage of initial body weight of $Agrp^{DTR/+}; Slc17a6^{lox/lox}$ mice after bilateral injection of AAV1-CreGFP virus ($n = 14$) or vehicle ($n = 10$) into the PBN. Asterisk, $P < 0.01$; ANOVA between mice with precise viral injection ($n = 7$) and vehicle injection ($n = 10$). Results are shown as means \pm s.e.m. Daggers indicate mice that were removed from experiment when they either lost more than 20% of body weight or seemed moribund.

suppression of glutamatergic signalling by the PBN should also protect against starvation when AgRP neurons are ablated. We used the $Slc17a6^{lox/lox}; Agrp^{DTR/+}$ mice again for this experiment and injected AAV1-CreGFP into the lateral PBN 9 days before DT treatment (Supplementary Fig. 5). The virally injected mice lost body weight during the first week after DT treatment but then gradually recovered, whereas the controls did not recover (Fig. 3c and Supplementary Fig. 8). In another cohort of mice, AAV1-CreGFP was injected into the PBN of $Slc17a6^{lox/lox}; Agrp^{DTR/+}$ mice, but they were not treated with DT. Those mice gained about 20% in body weight which was accompanied by about 16% increase in food intake over the next 3 weeks (Fig. 3d and Supplementary Fig. 9). Our results indicate that enhanced glutamatergic signalling by the PBN inhibits feeding and promotes weight loss, whereas lowering the glutamatergic output of the PBN promotes weight gain through an increase in feeding, possibly combined with a decrease in energy expenditure.

Our studies reveal six manipulations that allow mice to survive after the ablation of AgRP neurons: enhancement of GABA_A receptor signalling in the PBN with bretazenil⁵; suppression of 5-HT₃ receptor signalling in the NTS with ondansetron; disablement of serotonergic input to the NTS through viral-mediated removal of Tph2; reduction of glutamatergic signalling by the NTS by the removal of Vglut2; reduction of glutamatergic activation of the PBN by decreasing the number of NMDA receptors; or reduction of efferent glutamatergic

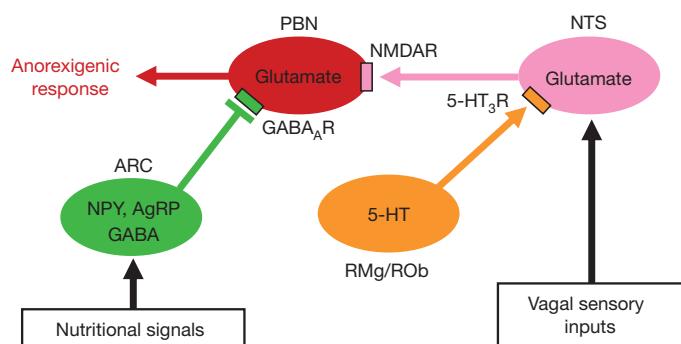


Figure 4 | Diagram illustrating circuitry that mediates loss of appetite after acute ablation of hypothalamic AgRP neurons. AgRP neurons co-expressing AgRP, neuropeptide Y (NPY) and GABA send inhibitory projections to the PBN. Serotonergic neurons residing in the RMg and ROb inhibit feeding through excitation of postsynaptic neurons in rostral NTS that express 5-HT₃ receptors (5-HT₃R). This subpopulation of NTS neurons, by integration of visceral and gustatory inputs, sends excitatory glutamate signalling to the lateral PBN neurons that express NMDA receptors (NMDAR). Nutritional signalling from the hypothalamus and sensory signals may interact within the PBN to promote appropriate feeding responses. ARC, arcuate nucleus; GABA_AR, GABA_A receptor.

signals from the PBN. These observations support the circuit depicted in Fig. 4. We suggest that a subpopulation of neurons in the PBN integrates visceral and gustatory information from the NTS with energy-balance signals emanating from AgRP neurons. The NTS responds to vagal inputs as well as gut-derived hormones, while AgRP neurons detect nutrient levels and respond to hormones such as insulin, leptin and ghrelin^{17–19}. Consequently, the appetitive response can be modulated by food palatability and visceral condition in a manner dictated by current energy balance.

Some studies have suggested that serotonin exerts its anorectic effects by differential actions on 5-HT_{1b} and 5-HT_{2c} receptors in the hypothalamus to stimulate melanocortin signalling^{20–22}, whereas a recent study indicated that serotonergic neurons in ventral raphe nuclei respond to food restriction by an elevated Fos signal²³. We show here that serotonin from the ROb and RMg acts on 5-HT₃ receptors in the NTS to inhibit feeding after the ablation of AgRP neurons; thus, some of the anorectic effects of serotonin reuptake inhibitors, such as fenfluramine, may also be mediated in the brainstem. Classical mapping studies reveal projections from the PBN to the amygdala, thalamus, hypothalamus and other brain regions²⁴. Further characterization and manipulation of PBN circuits that control feeding will be greatly facilitated by the identification of genes that are expressed exclusively by the relevant subpopulations of PBN neurons. Our experiments help define an important neural pathway within which some unique therapeutic targets have been characterized that could be valuable for the development of new treatments of various eating disorders, including nausea and anorexia nervosa^{25,26}.

METHODS SUMMARY

AgRP^{DTR/+} mice, *Grin1*^{lox/lox} mice and *Slc17a6*^{lox/lox} mice on a C57Bl/6 background were generated and identified genetically by PCR of tail DNA as described previously^{4,27,28}. A conditional *Tph2* targeting construct was prepared by flanking the first exon with *loxP* sites; the generation of *Tph2*^{lox/lox} mice on a mixed C57Bl/6 x 129/Sv background was based on standard protocols. All experiments were performed with male mice about 8 weeks old. To ablate AgRP neurons, mice carrying the *AgRP*^{DTR} allele were injected twice intramuscularly with DT (50 µg kg⁻¹; List Biological Laboratories, Campbell); the injections were 2 days apart⁴. The extent of ablation (more than 95%) was determined by immunohistochemistry^{4–7}. Production of adeno-associated virus 1 (AAV1-CreGFP) and canine adenovirus 2 (CAV2-Cre) followed the protocols described previously^{29,30}. For injection of virus into the NTS or PBN, mice were anaesthetized and virus was injected bilaterally or unilaterally through a 5-µl Hamilton syringe. For drug treatments, Alzet 14-day minipumps (model 1002) loaded with 100 µl of ondansetron (6 mg ml⁻¹; Sigma-Aldrich) were implanted subcutaneously on the back of anaesthetized mice 4 days before DT treatment. Alternatively, cannulas (28 gauge; Plastics One) were placed into fourth ventricles and connected by tubing with the subcutaneous minipumps that were loaded with ondansetron (0.6 mg ml⁻¹). For feeding assays, mice were transferred to BioDAQ Food and Water Intake Monitor (Research Diets) supplied with water and low-fat chow diet (D12450B). The mice were allowed to acclimatize for 3 days before the initiation of each experiment and data collection. Body weight and total food intake were recorded every 24 h. Feeding and drinking activity was recorded in accordance with the manufacturer's suggested protocol.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 July 2011; accepted 24 January 2012.

Published online 14 March 2012.

1. Wu, Q. & Palmiter, R. D. GABAergic signaling by AgRP neurons prevents anorexia via a melanocortin-independent mechanism. *Eur. J. Pharmacol.* **660**, 21–27 (2011).
2. Morton, G. J., Cummings, D. E., Baskin, D. G., Barsh, G. S. & Schwartz, M. W. Central nervous system control of food intake and body weight. *Nature* **443**, 289–295 (2006).
3. Gropp, E. et al. Agouti-related peptide-expressing neurons are mandatory for feeding. *Nature Neurosci.* **8**, 1289–1291 (2005).
4. Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310**, 683–685 (2005).

5. Wu, Q., Boyle, M. P. & Palmiter, R. D. Loss of GABAergic signaling by AgRP neurons to the parabrachial nucleus leads to starvation. *Cell* **137**, 1225–1234 (2009).
6. Wu, Q., Howell, M. P. & Palmiter, R. D. Ablation of neurons expressing agouti-related protein activates Fos and gliosis in postsynaptic target regions. *J. Neurosci.* **28**, 9218–9226 (2008).
7. Wu, Q., Howell, M. P., Cowley, M. A. & Palmiter, R. D. Starvation after AgRP neuron ablation is independent of melanocortin signaling. *Proc. Natl. Acad. Sci. USA* **105**, 2687–2692 (2008).
8. Swank, M. W. & Bernstein, I. L. c-Fos induction in response to a conditioned stimulus after single trial taste aversion learning. *Brain Res.* **636**, 202–208 (1994).
9. Yamamoto, T. Neural substrates for the processing of cognitive and affective aspects of taste in the brain. *Arch. Histol. Cytol.* **69**, 243–255 (2006).
10. Berridge, K. C. & Pecina, S. Benzodiazepines, appetite, and taste palatability. *Neurosci. Biobehav. Rev.* **19**, 121–131 (1995).
11. Gershon, M. D. & Tack, J. The serotonin signaling system: from basic understanding to drug development for functional GI disorders. *Gastroenterology* **132**, 397–414 (2007).
12. Barnes, N. M., Hales, T. G., Lummis, S. C. & Peters, J. A. The 5-HT₃ receptor—the relationship between structure and function. *Neuropharmacology* **56**, 273–284 (2009).
13. Herbert, H., Moga, M. M. & Saper, C. B. Connections of the parabrachial nucleus with the nucleus of the solitary tract and the medullary reticular formation in the rat. *J. Comp. Neurol.* **293**, 540–580 (1990).
14. Jhamandas, J. H. & Harris, K. H. Excitatory amino acids may mediate nucleus tractus solitarius input to rat parabrachial neurons. *Am. J. Physiol.* **263**, R324–R330 (1992).
15. Walther, D. J. & Bader, M. A unique central tryptophan hydroxylase isoform. *Biochem. Pharmacol.* **66**, 1673–1680 (2003).
16. Thor, K. B. & Heike, C. J. Serotonin- and substance P-containing projections to the nucleus tractus solitarius of the rat. *J. Comp. Neurol.* **265**, 275–293 (1987).
17. Abizaid, A. & Horvath, T. L. Brain circuits regulating energy homeostasis. *Regul. Pept.* **149**, 3–10 (2008).
18. Grill, H. J. Distributed neural control of energy balance: contributions from hindbrain and hypothalamus. *Obesity (Silver Spring)* **14** (Suppl 5), 216S–221S (2006).
19. Berthoud, H. R. & Morrison, C. The brain, appetite, and obesity. *Annu. Rev. Psychol.* **59**, 55–92 (2008).
20. Heisler, L. K. et al. Serotonin reciprocally regulates melanocortin neurons to modulate food intake. *Neuron* **51**, 239–249 (2006).
21. Xu, Y. et al. 5-HT_{2c}Rs expressed by pro-opiomelanocortin neurons regulate energy homeostasis. *Neuron* **60**, 582–589 (2008).
22. Xu, Y. et al. A serotonin and melanocortin circuit mediates d-fenfluramine anorexia. *J. Neurosci.* **30**, 14630–14634 (2010).
23. Takase, L. F. & Nogueira, M. I. Patterns of fos activation in rat raphe nuclei during feeding behavior. *Brain Res.* **1200**, 10–18 (2008).
24. Fulwiler, C. E. & Saper, C. B. Subnuclear organization of the efferent connections of the parabrachial nucleus in the rat. *Brain Res.* **319**, 229–259 (1984).
25. Rask-Andersen, M., Olszewski, P. K., Levine, A. S. & Schioto, H. B. Molecular mechanisms underlying anorexia nervosa: focus on human gene association studies and systems controlling food intake. *Brain Res. Brain Res. Rev.* **62**, 147–164 (2010).
26. Kaye, W. Neurobiology of anorexia and bulimia nervosa. *Physiol. Behav.* **94**, 121–135 (2008).
27. Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327–1338 (1996).
28. Hnasko, T. S. et al. Vesicular glutamate transport promotes dopamine storage and glutamate corelease *in vivo*. *Neuron* **65**, 643–656 (2010).
29. Kremer, E. J., Boutin, S., Chillon, M. & Danos, O. Canine adenovirus vectors: an alternative for adenovirus-mediated gene transfer. *J. Virol.* **74**, 505–512 (2000).
30. Kaplitt, M. G. et al. Long-term gene expression and phenotypic correction using adeno-associated virus vectors in the mammalian brain. *Nature Genet.* **8**, 148–154 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Froelick, J. Wang and K. Battani for help with histology; A. Rainwater for help with mouse breeding; A. Quintana for propagating CAV2-Cre virus and preparing AAV1-CreGFP virus; and A. Guler and M. Carter for helpful comments on the manuscript. This work was supported in part by National Institutes of Health grant DA024908 to R.D.P.

Author Contributions Q.W. and R.D.P. designed the research. Q.W. performed experiments and analysed the data. M.C. provided the conditional *Tph2* mouse line. R.D.P. and Q.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.D.P. (palmiter@uw.edu).

METHODS

Animal maintenance and neuron ablation. Mice were housed in a temperature- and humidity-controlled facility with a 12-h light cycle. All animal care and experimental procedures were approved by the Institutional Animal Care and Use Committee at the University of Washington. In compliance with our approved protocol, all experiments were terminated if the body weight of mice fell to 80% of their original body weight or they seemed moribund. *Agrp*^{DTR/+} mice, *Grin1*^{lox/lox} mice and *Slc17a6*^{lox/lox} mice were generated and genetically identified by PCR of tail DNA as described previously^{4,27,28}. A conditional *Tph2* targeting construct was prepared by flanking the first exon with *loxP* sites along with a *frt*-flanked *SV-Neo* gene in the first intron. The construct was electroporated in G4 ES cells, and correctly targeted clones were identified by Southern blotting. After removal of the *SV-Neo* gene by breeding with a mouse expressing FLP recombinase, heterozygotes were bred to generate *Tph2*^{lox/lox} mice that were used for viral injection. Details are available from the authors on request.

All except *Tph2*^{lox/lox} mice were on the C57Bl/6 background (more than nine generations backcrossed); *Tph2*^{lox/lox} mice were on a mixed 129/Sv × C57Bl/6 background. *Agrp*^{DTR/DTR} male mice were bred with *Slc17a6*^{lox/lox} female mice to generate *Agrp*^{DTR/+}; *Slc17a6*^{lox/+} mice, which were further bred to each other to create *Agrp*^{DTR/+}; *Slc17a6*^{lox/lox} mice and *Agrp*^{DTR/+}; *Slc17a6*^{lox/+} control mice. A similar breeding strategy was adopted when generating *Agrp*^{DTR/+}; *Tph2*^{lox/lox} mice and *Agrp*^{DTR/+}; *Grin1*^{lox/lox} mice. Mice were group housed with a standard chow diet (LabDiet 5053) and water provided *ad libitum* until the beginning of the experiments. All experiments were performed with male mice about 8 weeks old. To ablate AgRP neurons, mice carrying the *Agrp*^{DTR} allele were twice injected intramuscularly with DT (50 µg kg⁻¹; List Biological Laboratories, Campbell); the injections were 2 days apart⁴. The extent of ablation (more than 95%) was determined by immunohistochemistry⁴⁻⁷.

Viral injections. Production of adeno-associated virus 1 (AAV1-CreGFP) and canine adenovirus 2 (CAV2-Cre) followed the protocols described previously^{29,30}. For injection of CAV2-Cre or AAV1-CreGFP virus into the rostral NTS, mice were anaesthetized and virus (or PBS as the vehicle) was injected bilaterally or unilaterally (1 µl of roughly 10¹⁰ particles per µl each side) through a Hamilton syringe (size 5 µl; Hamilton, Reno), using stereotactic coordinates ±0.8 mm (x axis), -7.1 mm (y axis) and -4.3 mm (z axis). Similarly, AAV1-CreGFP virus (or PBS as the vehicle) was injected into the PBN bilaterally or unilaterally, using stereotactic coordinates ±1.0 mm (x axis), -5.3 mm (y axis) and -3.3 mm (z axis). Brain samples from all mice were collected at the end of the behavioural experiment and processed for immunohistological analysis. For all viral injection experiments, only a fraction (indicated in figure legends) of the DT-treated mice survived AgRP-neuron ablation; subsequent evaluation of CreGFP expression revealed that failure to rescue was associated with poor placement or inadequate viral transduction.

Drug treatments. Alzet 14-day minipumps (model 1002; Durect, Cupertino) loaded with 100 µl of ondansetron (6 mg ml⁻¹ in saline; Sigma-Aldrich, St Louis) were implanted subcutaneously on the back of anaesthetized mice 4 days before DT treatment. These minipumps dispense 0.25 µl h⁻¹. Alternatively, cannulas (28 gauge; Plastics One, Roanoke) were placed into fourth ventricles

under anaesthesia and the subcutaneous minipumps, which were loaded with ondansetron (0.6 mg ml⁻¹ in saline), were connected to the cannulas by tubing (PE60; Stoelting, Wood Dale) that was threaded under the skin to help prevent the mice from dislodging it. For some experiments, the minipumps (ondansetron, 0.6 mg ml⁻¹) were connected to bilateral cannulas (28 gauge; Plastics One) directed to specific brain regions by using the following stereotactic coordinates: PBN, ±1 mm (x axis), -5.3 mm (y axis) and -3.3 mm (z axis); NTS, ±0.8 mm (x axis), -7.1 mm (y axis) and -4.3 mm (z axis). The patency and placement of the bilateral minipump were verified at the end of each experiment, and brain samples were processed for immunohistological analysis.

Food intake and body weight measurements. For feeding assays, mice were transferred to BioDAQ Food and Water Intake Monitor (Research Diets, New Brunswick) supplied with water and low-fat (3.85 kcal ml⁻¹) chow diet (D12450B; Research Diets). The mice were allowed to acclimatize for 3 days before initiation of each experiment and data collection. Body weight and total food intake were recorded every 24 h. Feeding and drinking activity was recorded in accordance with the manufacturer's suggested protocol.

Immunohistochemistry. Mice were killed by CO₂ asphyxiation and perfused transcardially with ice-cold PBS buffer containing 4% paraformaldehyde. Brains were dissected and postfixed overnight at 4 °C in the fixation buffer. Free-floating brain sections (25 µm) were washed three times in PBS containing 0.1% Triton X-100 (PBST buffer) solution (15 min each wash) and then blocked with 3% normal donkey serum in PBST for 2–3 h at about 23 °C. Rabbit anti-AgRP (dilution 1:1,500; Phoenix Pharmaceuticals, Belmont), rabbit anti-Fos (dilution 1:1,000; Millipore, Temecula), monoclonal anti-tryptophan hydroxylase (dilution 1:1,500; Sigma-Aldrich) and rabbit anti-dopa decarboxylase (equivalent to AADC; dilution 1:500; Millipore) were applied to the sections for incubation overnight at 4 °C, followed by three 15-min rinses in PBST. Finally, sections were incubated in Cy2- or Cy3-labelled secondary antibody (dilution 1:300; Jackson Immunolaboratory, West Grove) before visualization. Images were captured with a digital camera mounted on a Leica TCS SP1 confocal microscope (Leica Microsystems); all paired photos were obtained through the same system settings. For each group of mice, at least eight sections from four different mice were analysed.

Data analyses. Quantification of *Tph2*-positive and AADC-positive cells was performed with the NIH Image software (National Institutes of Health). Anatomical correlations of brain sections and delineation of individual nuclei were determined by comparing landmarks of Nissl staining images with those given in the stereotactic atlas. From the anatomically matched sections, a region of interest of the same size was further defined. Meanwhile, an optimized threshold that can discern round nuclei from partly stained ones as well as background noise was preset for all measurements. For all experiments only those mice with correct placement of cannula or viral injections were compared with the control group. Unless otherwise stated, data sets collected from all experiments were analysed by one-way ANOVA followed by the Student–Newman–Keuls method for statistical significance; results were plotted as means ± s.e.m. Post-hoc analysis was performed when group differences were significant by ANOVA at *P* < 0.05.

Visualizing molecular juggling within a B₁₂-dependent methyltransferase complex

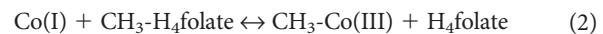
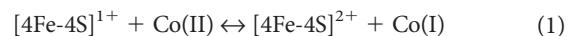
Yan Kung¹, Nozomi Ando^{1,2}, Tzanko I. Doukov¹†, Leah C. Blasiak¹†, Güneş Bender³, Javier Seravalli⁴, Stephen W. Ragsdale³ & Catherine L. Drennan^{1,2,5}

Derivatives of vitamin B₁₂ are used in methyl group transfer in biological processes as diverse as methionine synthesis in humans and CO₂ fixation in acetogenic bacteria^{1–3}. This seemingly straightforward reaction requires large, multimodular enzyme complexes that adopt multiple conformations to alternately activate, protect and perform catalysis on the reactive B₁₂ cofactor. Crystal structures determined thus far have provided structural information for only fragments of these complexes^{4–12}, inspiring speculation about the overall protein assembly and conformational movements inherent to activity. Here we present X-ray crystal structures of a complete 220 kDa complex that contains all enzymes responsible for B₁₂-dependent methyl transfer, namely the corrinoid iron-sulphur protein and its methyltransferase from the model acetogen *Moorella thermoacetica*. These structures provide the first three-dimensional depiction of all protein modules required for the activation, protection and catalytic steps of B₁₂-dependent methyl transfer. In addition, the structures capture B₁₂ at multiple locations between its ‘resting’ and catalytic positions, allowing visualization of the dramatic protein rearrangements that enable methyl transfer and identification of the trajectory for B₁₂ movement within the large enzyme scaffold. The structures are also presented alongside *in crystallo* spectroscopic data, which confirm enzymatic activity within crystals and demonstrate the largest known conformational movements of proteins in a crystalline state. Taken together, this work provides a model for the molecular juggling that accompanies turnover and helps explain why such an elaborate protein framework is required for such a simple, yet biologically essential reaction.

B₁₂-dependent methyl transfer lies at the heart of methylation biochemistry and is an essential reaction in human health and microbial CO₂ sequestration^{2,3}. In humans, methionine synthase (MetH) methylates homocysteine to form methionine to maintain cellular pools of folate (vitamin B₉) and S-adenosylmethionine (AdoMet), the universal methyl donor. MetH mutation or vitamin B₁₂ deficiency can cause serious health consequences, including megaloblastic anaemia and birth abnormalities such as neural tube defects¹³. Acetogenic bacteria, including *M. thermoacetica*, use the corrinoid iron-sulphur protein (CFeSP) and its methyltransferase (MeTr) together to catalyse methyl transfer in the Wood-Ljungdahl carbon fixation pathway for growth on CO₂ as the sole carbon source¹⁴.

For both MetH and CFeSP/MeTr, methyltetrahydrofolate (CH₃-H₄folate) is the methyl donor, and a protein-bound B₁₂ derivative (cobalamin for MetH and 5'-methoxybenzimidazolyl cobamide for CFeSP) is the methyl carrier. In acetogenic bacteria, the CH₃-H₄folate methyl group is derived from enzymatic reduction of CO₂, whereas in humans, CH₃-H₄folate is the predominant circulating form of the vitamin. Although CH₃-H₄folate is the common methyl source, methyl removal from the N⁵ tertiary amine is chemically challenging because the product, tetrahydrofolate (H₄folate), is a poor leaving

group¹. Therefore, a particularly powerful nucleophile is required, and B₁₂ with cobalt in the +1 oxidation state, a Co(I) species dubbed a ‘supernucleophile’¹⁵, is recruited. Such strong reactivity comes at a price: reducing the inactive Co(II) state to active Co(I) is thermodynamically challenging, as the Co(II/I) reduction potential is one of the lowest in nature, −504 mV in CFeSP and −526 mV in MetH^{16,17}. In CFeSP, an electron is first delivered from a partner protein to an Fe₄S₄ cluster harboured by an activation domain^{18,19}. The electron is then passed to Co(II) to yield Co(I) (equation (1)), which attacks CH₃-H₄folate to form CH₃-Co(III) (equation (2)). CFeSP then delivers the methyl group to the Ni₂Fe₄S₄ active site metallocluster (A-cluster) of acetyl-CoA synthase (ACS), where it becomes the methyl of acetyl-CoA, and B₁₂ returns to its nucleophilic Co(I) state.



During the catalytic cycle of both MetH and CFeSP/MeTr (Supplementary Fig. 1), a series of ‘molecular juggling’ acts must be performed in which domains rearrange to contact the B₁₂ cofactor. Crystal structures of a MetH B₁₂-binding fragment⁴ and CFeSP from *Carboxydotothermus hydrogenoformans* (ChCFeSP)⁹ both depict a ‘resting’ state, where B₁₂ is buried by a protective ‘capping’ domain, shielded from unwanted chemistry but inaccessible to substrate. Because methyl transfer uses S_N2 substitution²⁰, large conformational changes must ‘uncap’ B₁₂ before chemistry can occur. B₁₂ is ‘uncapped’ in structures of MetH fragments that depict B₁₂ activation^{7,11,12}, but no structure has been solved that shows B₁₂- and CH₃-H₄folate-binding domains together to illustrate methyl transfer.

To visualize this elusive methyl transfer complex, we determined a 2.38 Å resolution structure of folate-free CFeSP/MeTr from *M. thermoacetica* (Fig. 1 and Supplementary Table 1). The homodimeric MeTr component (58 kDa) is virtually identical to previous structures of both MeTr^{6,10} (Supplementary Fig. 2a), root mean squared deviation (r.m.s.d.) for C_α atoms 0.39 Å, and the analogous MetH domain that binds CH₃-H₄folate⁸, r.m.s.d. 1.03–1.08 Å. MeTr and MetH both use (β/α)₈ triosephosphate isomerase (TIM) barrels to bind and activate CH₃-H₄folate for nucleophilic attack. Two CFeSPs are present in the complex, each containing two subunits. The small subunit (35 kDa) is a TIM barrel which acts as the B₁₂ ‘cap’ in the ChCFeSP structure⁹, while the large subunit (48 kDa) has three domains joined by linkers: an amino (N)-terminal Fe₄S₄ activation domain (residues 1–57), a TIM barrel domain (residues 93–312) and a carboxy (C)-terminal B₁₂-binding domain (residues 325–446). With the exception of the Fe₄S₄ and B₁₂ domains, discussed below, both CFeSP copies align well to the ChCFeSP structure, r.m.s.d. 0.81–0.85 Å (Supplementary Fig. 2b).

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁴Department of Biochemistry, University of Nebraska, Lincoln, Nebraska 68588, USA. ⁵Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present addresses: Stanford Synchrotron Radiation Lightsource, Menlo Park, California 94025, USA (T.I.D.); Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore 21202, USA (L.C.B.).

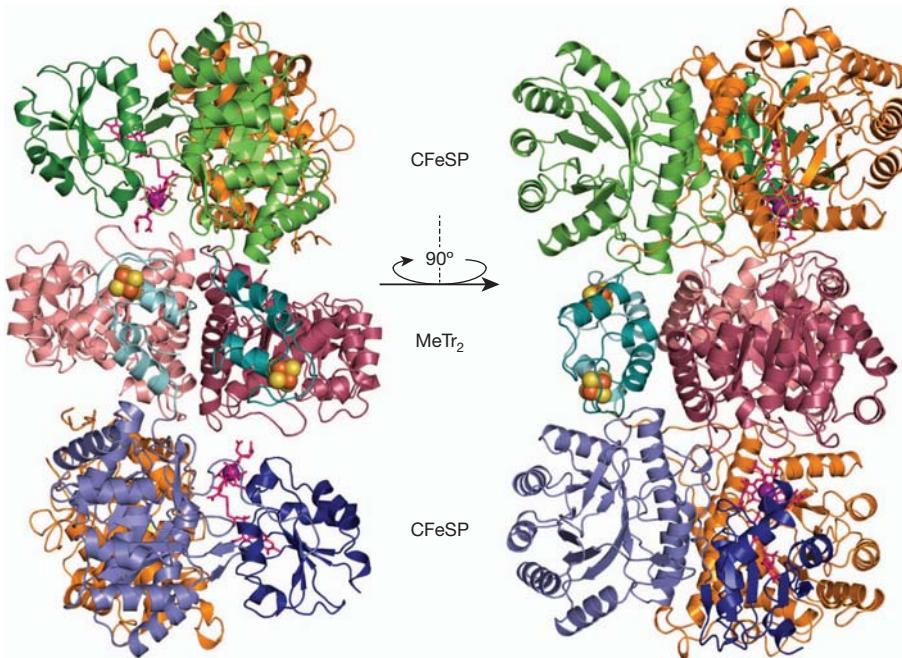


Figure 1 | The overall CFeSP/MeTr complex. Ribbon representation of MeTr homodimer (MeTr₂) in light and dark pink, CFeSP small subunits in orange, CFeSP large subunit Fe₄S₄ domains in teal and cyan, TIM barrel domains in

green and blue, and B₁₂ domains in dark green and dark blue. B₁₂ cofactors in magenta sticks with cobalt as violet spheres. Fe₄S₄ clusters in spheres: Fe in orange, S in yellow.

In the 220 kDa CFeSP/MeTr assembly (Fig. 1), the MeTr homodimer lies in the centre, with one CFeSP bound on either side. Each MeTr monomer has a C-terminal α -helix (residues 255–262) protruding from the TIM barrel rim. Contacts between this helix and its preceding loop with a CFeSP small subunit helix (residues 191–204) form the primary interactions between MeTr and CFeSP (Supplementary Fig. 3). Weak interactions between MeTr and CFeSP Fe₄S₄ domains have stabilized these highly flexible⁹ domains responsible for B₁₂ activation, allowing their visualization as bundles of short α -helices connected by long loops that coordinate the Fe₄S₄ cubane (Fig. 1 and Supplementary Figs 4 and 5). The Fe₄S₄ domains are observed to adopt a variety of positions that are all too far from the B₁₂ to afford reductive activation⁹. However, the long and primarily unstructured protein linkers that connect both the Fe₄S₄ and B₁₂ domains to the central TIM barrel must allow for the requisite flexibility for B₁₂ activation (Supplementary Figs 2b and 6).

B₁₂ domains of both MetH and CFeSP adopt Rossmann-like architectures that bind B₁₂ in the base-off conformation (Fig. 1)^{4,9}. High *B*-factors support the notion of flexibility mentioned above (Supplementary Fig. 7 and Supplementary Table 2), where electron density for both B₁₂ domains represents a highest occupancy position within an ensemble, rather than a sole conformation. In both CFeSPs, the average B₁₂ domain position resides between the ‘capping’ small subunit TIM barrel and the TIM barrel of a MeTr monomer, which are adjacent and nearly perpendicular to each other (Fig. 1). On average, the B₁₂ Co has shifted approximately 6.5 Å away from its ‘resting’ location towards the MeTr folate-binding site. B₁₂ in this structure is thus positioned ‘en route’ towards catalysis, with approximately 18 Å remaining to the methyl group of folate modelled into the MeTr active site, based on an alignment with the folate-bound MeTr structure¹⁰. In transitioning between ‘resting’ and ‘en route’ positions, the B₁₂ corrin ring breaks three interactions with the ‘capping’ domain and forms new contacts, including an H-bond with Asn 203 of MeTr (Fig. 2a, b).

Given the flexibility suggested by this structural analysis, we explored whether the B₁₂ domain can sample the 18 Å necessary to afford turnover within intact crystals, using anaerobic *in crystallo* ultraviolet-visible absorption spectroscopy to monitor the state of B₁₂. *In crystallo* and analogous solution spectra were collected in

parallel (Fig. 3) for the as-isolated Co(II) form of B₁₂. Reduction to Co(I) and methylation to CH₃-Co(III) were then achieved *in crystallo* and in solution, with all spectra matching well-established CFeSP absorption features^{18,19,21–23}. Importantly, these features disappear when light is passed through the solution surrounding the crystals, indicating that spectra represent protein in crystals and not protein that may have been liberated into the solution. Collectively, these data demonstrate enzymatic transfer of the CH₃-H₄folate methyl group to CFeSP-bound B₁₂, evidence that the B₁₂ domain is able to move at least 18 Å to trigger methyl transfer within the crystal. To our knowledge, this conformational movement represents the largest observed in a crystallized protein (Supplementary Discussion). Such dramatic B₁₂ domain movement is probably facilitated by the fact that CFeSP/MeTr is mostly composed of rigid TIM barrels that provide all the lattice contacts (Supplementary Fig. 8). Although their biosynthesis is energetically expensive, these high molecular mass TIM barrel scaffolds may be important for B₁₂-dependent methyltransferases to maintain structural integrity during the conformational gymnastics that alternately enable activation, protection and catalysis of the highly reactive B₁₂ cofactor. Thus, despite the small size of the transferred methyl moiety, these large conformational changes appear to necessitate large enzyme sizes.

Although the folate-free CFeSP/MeTr structure describes large B₁₂ domain movements that ‘uncap’ B₁₂ from the small subunit, it is interesting to consider why binding of CFeSP to MeTr does not simply position the B₁₂ domain directly over the MeTr active site. One explanation is that the structure represents an inactive complex; however, *in crystallo* results clearly demonstrate that CFeSP/MeTr crystals are active. Another explanation posits that an ensemble of ‘en route’ conformations exists when CH₃-H₄folate is absent, and that CH₃-H₄folate binding would shift the conformational equilibrium, moving B₁₂ closer to the folate-binding site. To obtain experimental support for this hypothesis, we solved additional CFeSP/MeTr structures co-crystallized with CH₃-H₄folate, with and without Ti(III) citrate as a reductant at 3.03 Å and 3.50 Å resolution, respectively. Absorption spectroscopy performed on these crystals shows that these structures represent a substrate form (CH₃-H₄folate bound, B₁₂ in the Co(II) state) and a product form (H₄folate bound, B₁₂ in the CH₃-Co(III) state) of the complex (Supplementary Fig. 9). Compared with

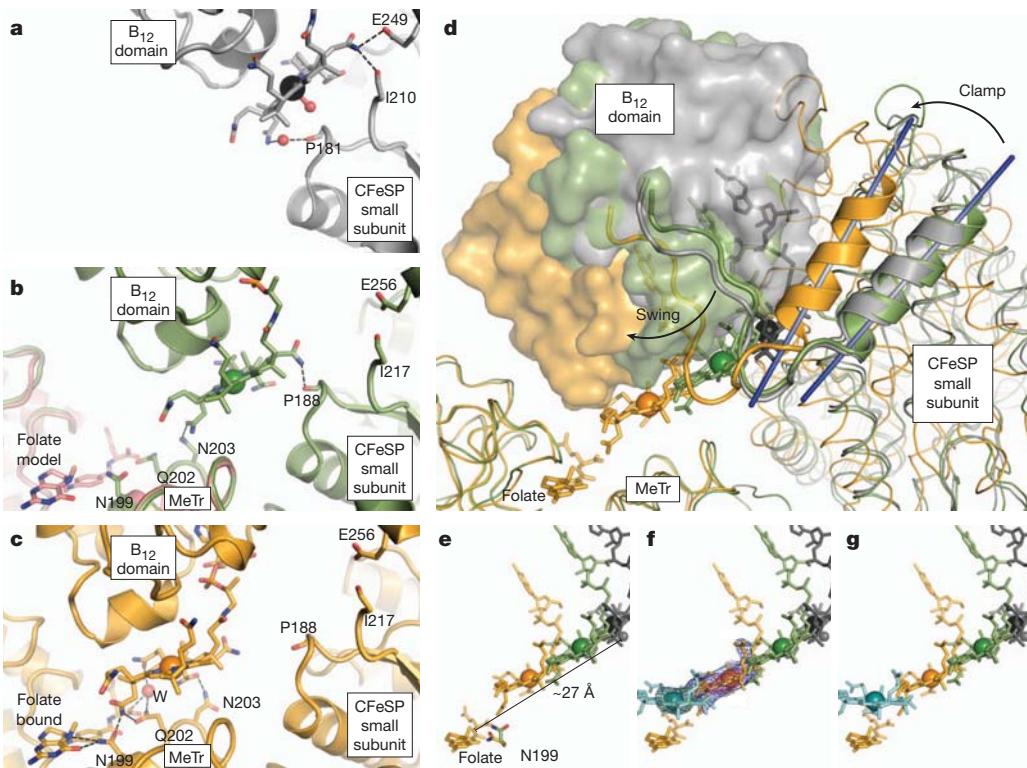


Figure 2 | Comparison of B₁₂ positions in 'resting' ChCFeSP, folate-free and folate-bound CFeSP/MeTr. **a**, ChCFeSP (grey ribbons, Co of B₁₂: black sphere). **b**, Folate-free CFeSP/MeTr (green ribbons, Co of B₁₂: green sphere) superimposed with CH₃-H₄folate-bound MeTr (Protein Data Bank accession number 2E7F, pink ribbons). **c**, Folate-bound CFeSP/MeTr (orange ribbons, Co of B₁₂: orange sphere). Parts **a**–**c** are identical in orientation; B₁₂ sticks coloured as C, ribbon colour; O, red; N, blue; P, orange. **d**, Superposition of ChCFeSP (grey), folate-free CFeSP/MeTr (green), and folate-bound CFeSP/MeTr

(orange) structures in **a**–**c**, highlighting one helix (thick ribbons) to show clamping motion (helix axes as straight blue lines) and B₁₂ (sticks) with 12-residue linker (thick ribbons) to B₁₂ domain (surface) to show swinging motion. **e**, Superposition of B₁₂ and CH₃-H₄folate in **d**, with Asn 199 shown for CFeSP/MeTr structures in sticks (C, ribbon colour; O, red; N, blue). **f**, Same as **e**, with 2F_o – F_c density in blue (1.0σ) and pink mesh (4.0σ), and F_o – F_c density in green mesh (3.0σ) for folate-bound CFeSP/MeTr structure. Putative alternative B₁₂ corrin: cyan. **g**, Superposition of B₁₂ cofactors and CH₃-H₄folate in **f**.

the folate-free structure, B₁₂ in both folate-bound structures has indeed moved even closer to the MeTr folate-binding site (by an average of 7.7 Å) and exhibits new H-bonding features (Fig. 2c). In these folate-bound structures, the B₁₂ corrin ring has severed all interactions with the 'capping' CFeSP small subunit and contacts only MeTr residues. Here, asparagine and glutamine residues that line the MeTr surface appear to participate in an 'amide hand off', sequentially passing B₁₂ along its trajectory as it progresses towards folate (Fig. 2b–c and Supplementary Fig. 10).

Interestingly, the terminal amide in this 'hand off', Asn 199, is strictly conserved in both MeTr and MetH and was previously shown to switch conformations between folate-free and folate-bound forms¹⁰, a feature also observed in the CFeSP/MeTr structures presented here (Fig. 2b, c, e). In apo-MeTr structures, Asn 199 points upwards and out of the active site, whereas in folate-bound MeTr structures Asn 199 turns down to H-bond with the N⁵ of folate. Because N199A mutation moderately hinders folate binding (20-fold in dissociation constant, K_d) but dramatically compromises catalytic efficiency (k_{cat}/K_m) by 25,000-fold, Asn 199 is thought to be important for formation of the transition state¹⁰. In our CFeSP/MeTr structures, we observe a new role for Asn 199 in B₁₂ domain conformational switching: when folate is absent, Asn 199 points out of the active site, blocking a closer B₁₂ position. However, when folate binds and Asn 199 reorients to H-bond with folate, space becomes available for B₁₂ to move closer to the MeTr folate-binding site. Therefore, the position of Asn 199 itself could help shift the conformational equilibrium of the B₁₂ domain, signalling that substrate has bound to MeTr. Asn 199 is an ideal signal for substrate binding, as it is the only MeTr residue known to reposition upon folate binding¹⁰.

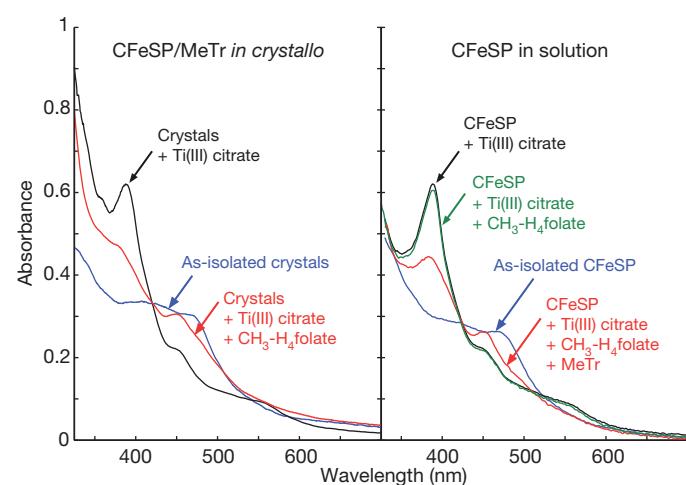


Figure 3 | Methyl transfer activity of CFeSP/MeTr crystals by ultraviolet-visible absorption spectroscopy. As-isolated spectra (blue lines) for CFeSP/MeTr crystals and CFeSP in solution similarly have broad features at approximately 400 and 470 nm arising from the Fe₄S₄ cluster and the B₁₂ corrin. Following established protocols^{18,19,22,23,26,27}, B₁₂ reduction was achieved with Ti(III) citrate, yielding a sharp 390 nm peak indicative of active Co(I) in both solution and *in crystallo* spectra (black lines). Further treatment with CH₃-H₄folate yields decreased absorbance at 390 nm and a new peak at 450 nm (red lines), characteristic of the product complex (protein-bound CH₃-Co(III)^{18,19,22,23}). A control reaction (green line) confirms that turnover does not occur from free CH₃-H₄folate without MeTr, and the 450 nm peak indicates that B₁₂ remains CFeSP-bound (free B₁₂ has a peak at approximately 520 nm instead^{18,21,27}).

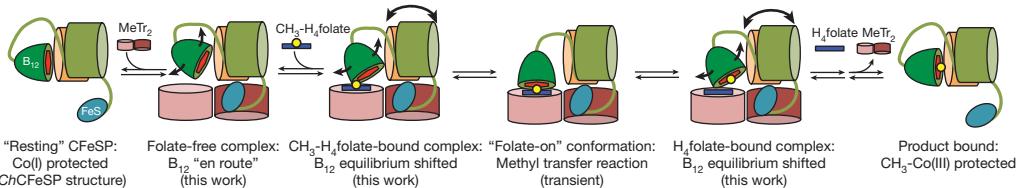


Figure 4 | Cartoon model of B_{12} -dependent methyl transfer in CFeSP/MeTr. For simplicity, only one of the two CFeSP heterodimers is shown. Protein domains are coloured as in Fig. 1, loops represent linkers, red hexagon

Displacement of the B_{12} domain from its resting position to the position nearest the folate-binding site can be attributed to two independent conformational changes within the complex, best described as 'swinging' and 'clamping' motions (Fig. 2d). The B_{12} domain can 'swing' relative to the rest of CFeSP (Supplementary Fig. 11), and CFeSP can 'clamp' the B_{12} domain towards the MeTr active site (Fig. 2d and Supplementary Fig. 12). Despite varying degrees of 'clamping' over a range of approximately 14° across the structures, the interface between the CFeSP small subunit and MeTr is preserved (Supplementary Fig. 13).

Although folate binding shifts the average position of the B_{12} domain closer to the MeTr folate-binding site, the B_{12} Co is still too far for $S_{N}2$ methyl transfer (Fig. 2e). Intriguingly, a large, continuous electron density peak is present in $2F_o - F_c$, $F_o - F_c$ and composite omit maps, emanating from the corrin ring and stretching directly over the folate-binding site, suggestive of an alternative, low-occupancy corrin conformation (Fig. 2f and Supplementary Fig. 14). A trial refinement of a putative corrin ring at 40% occupancy satisfies the $F_o - F_c$ difference maps (Supplementary Fig. 15), positioning B_{12} over the folate-binding site.

The multiple positions of B_{12} captured here (Fig. 2g) highlight the conformational flexibility of the CFeSP/MeTr scaffold and provide a framework to understand the molecular juggling of domains during B_{12} -dependent methyl transfer (Fig. 4). Before MeTr binding, the CFeSP B_{12} domain rests against the 'capping' small subunit, as in the structure of ChCFeSP⁹, with reactive Co(I) of B_{12} protected ('resting' state). From this conformation, either the 'cap' or the B_{12} domain must move to allow substrate access. Our folate-free CFeSP/MeTr structure indicates that upon MeTr binding, the B_{12} domain becomes 'loosened' and flexible, adopting an ensemble of conformations that lie en route towards the MeTr active site. Here, the reactive B_{12} species would be protected by the CFeSP small subunit and MeTr TIM barrels. $CH_3\text{-H}_4\text{folate}$ binding to MeTr accompanied by movement of Asn 199 shifts the equilibrium of B_{12} domain conformers, placing B_{12} closer to folate, as in our folate-bound CFeSP/MeTr structures, with B_{12} protection afforded by MeTr. It is notable that even after $CH_3\text{-H}_4\text{folate}$ binds, the major B_{12} position is still not directly over the folate methyl group, as such a position is expected to be transient. After methyl transfer, the B_{12} domain can return to the small subunit to 're-cap' the methylated B_{12} product, protected by the small subunit TIM barrel.

Overall, our data indicate that B_{12} domain movement is not a simple, two-state switch between 'resting' and 'catalytic' conformations. Instead, a flexible B_{12} domain samples an ensemble of conformations, where subtle shifts of the conformational equilibrium place B_{12} progressively closer to the active site, thereby increasing the population of conformers capable of methyl transfer without obstructing substrate access or hindering domain movement. This model is consistent with MetH studies where ligation, alkylation and redox state of the B_{12} cobalt can favour/disfavour various binding modes, alternately shifting the equilibrium of conformers for ordered domain rearrangements during the reaction cycle^{11,12,24,25}. We further identify MeTr residues that contact B_{12} along its trajectory, ending with Asn 199. In MetH the B_{12} ligating residue His 759 has been shown to play a dual role in catalysis and in signalling conformational shifts¹¹. The strictly conserved, folate-binding Asn 199 of MeTr could similarly play a dual role in both catalysis and conformational signalling. We thus expect this

is B_{12} , blue rectangle is folate and transferred methyl group is shown as a yellow sphere. Curved arrows denote 'swinging' and 'clamping' motions.

model for dynamic domain juggling, communicated by residues involved in substrate and cofactor binding, to be a common theme in methyl transfer between the B vitamins folate and B_{12} .

METHODS SUMMARY

CFeSP and MeTr were expressed and purified anaerobically from *M. thermoacетica* ATCC 39073 and from recombinant *Escherichia coli*, respectively. Crystals were grown anaerobically by hanging drop vapour diffusion. Diffraction data were collected at 24ID-C at the Advanced Photon Source, Argonne National Laboratory, and 5.0.2 and 8.2.2 at the Advanced Light Source, Lawrence Berkeley National Laboratory. Structures were solved by molecular replacement. Data collection and refinement statistics are presented in Supplementary Table 1, and representative electron densities for protein domains and for cofactors/substrate are shown in Supplementary Figs 16–21. Solution and *in crystallo* absorption spectra were collected as described in the text and in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 December 2010; accepted 3 February 2012.

Published online 14 March 2012.

1. Matthews, R. G. Cobalamin-dependent methyltransferases. *Acc. Chem. Res.* **34**, 681–689 (2001).
2. Banerjee, R. B. & Ragsdale, S. W. The many faces of vitamin B_{12} : catalysis by cobalamin-dependent enzymes. *Annu. Rev. Biochem.* **72**, 209–247 (2003).
3. Matthews, R. G., Koutmos, M. & Datta, S. Cobalamin-dependent and cobamide-dependent methyltransferases. *Curr. Opin. Struct. Biol.* **18**, 658–666 (2008).
4. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. How a protein binds B_{12} : a 3.0 Å X-ray structure of B_{12} -binding domains of methionine synthase. *Science* **266**, 1669–1674 (1994).
5. Dixon, M. M., Huang, S., Matthews, R. G. & Ludwig, M. L. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B_{12} . *Structure* **4**, 1263–1275 (1996).
6. Doukov, T., Seravalli, J., Stezowski, J. J. & Ragsdale, S. W. Crystal structure of a methyltetrahydrofolate- and corrinoid-dependent methyltransferase. *Structure* **8**, 817–830 (2000).
7. Bandarian, V. *et al.* Domain alternation switches B_{12} -dependent methionine synthase to the active conformation. *Nature Struct. Biol.* **9**, 53–56 (2002).
8. Evans, J. C. *et al.* Structures of the N-terminal module imply large domain motions during catalysis by methionine synthase. *Proc. Natl. Acad. Sci. USA* **101**, 3729–3736 (2004).
9. Svetlichnaya, T., Svetlichnyi, V., Meyer, O. & Dobbek, H. Structural insights into methyltransfer reactions of a corrinoid iron-sulfur protein involved in acetyl-CoA synthesis. *Proc. Natl. Acad. Sci. USA* **103**, 14331–14336 (2006).
10. Doukov, T. I., Hemmi, H., Drennan, C. L. & Ragsdale, S. W. Structural and kinetic evidence for an extended hydrogen-bonding network in catalysis of methyl group transfer: role of an active site asparagine residue in activation of methyl transfer by methyltransferases. *J. Biol. Chem.* **282**, 6609–6618 (2007).
11. Datta, S., Koutmos, M., Partridge, K. A., Ludwig, M. L. & Matthews, R. G. A. Disulfide-stabilized conformer of methionine synthase reveals an unexpected role for the histidine ligand of the cobalamin cofactor. *Proc. Natl. Acad. Sci. USA* **105**, 4115–4120 (2008).
12. Koutmos, M., Datta, S., Partridge, K. A., Smith, J. L. & Matthews, R. G. Insights into the reactivation of cobalamin-dependent methionine synthase. *Proc. Natl. Acad. Sci. USA* **106**, 18527–18532 (2009).
13. Banerjee, R. B. & Matthews, R. G. Cobalamin-dependent methionine synthase. *FASEB J.* **4**, 1450–1459 (1990).
14. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. *Biochim. Biophys. Acta* **1784**, 1873–1898 (2008).
15. Schrauzer, G. N. & Deutsch, E. Reactions of cobalt(I) supernucleophiles. The alkylations of vitamin B_{12} s, cobaloximes(I), and related compounds. *J. Am. Chem. Soc.* **91**, 3341–3350 (1969).
16. Harder, S. R., Lu, W.-P., Feinberg, B. A. & Ragsdale, S. W. Spectroelectrochemical studies of the corrinoid/iron-sulfur protein involved in acetyl coenzyme a synthesis by *Clostridium thermoaceticum*. *Biochemistry* **28**, 9080–9087 (1989).
17. Banerjee, R. B., Harder, S. R., Ragsdale, S. W. & Matthews, R. G. Mechanism of reductive activation of cobalamin-dependent methionine synthase: an electron

- paramagnetic resonance spectroelectrochemical study. *Biochemistry* **29**, 1129–1135 (1990).
18. Menon, S. & Ragsdale, S. W. Role of the [4Fe-4S] cluster in reductive activation of the cobalt center of the corrinoid iron–sulfur protein from *Clostridium thermoaceticum* during acetate biosynthesis. *Biochemistry* **37**, 5689–5698 (1998).
 19. Menon, S. & Ragsdale, S. W. The role of an iron–sulfur cluster in an enzymatic methylation reaction. *J. Biol. Chem.* **274**, 11513–11518 (1999).
 20. Zydowsky, T. M. *et al.* Stereochemical analysis of the methyl transfer catalyzed by cobalamin-dependent methionine synthase from *Escherichia coli* B. *J. Am. Chem. Soc.* **108**, 3152–3153 (1986).
 21. Ragsdale, S. W., Lindahl, P. A. & Münck, E. Mössbauer, EPR, and optical studies of the corrinoid/iron–sulfur protein involved in the synthesis of acetyl coenzyme A by *Clostridium thermoaceticum*. *J. Biol. Chem.* **262**, 14289–14297 (1987).
 22. Zhao, S., Roberts, D. L. & Ragsdale, S. W. Mechanistic studies of the methyltransferase from *Clostridium thermoaceticum*: origin of the pH dependence of the methyl group transfer from methyltetrahydrofolate to the corrinoid/iron–sulfur protein. *Biochemistry* **34**, 15075–15083 (1995).
 23. Seravalli, J., Zhao, S. & Ragsdale, S. W. Mechanism of transfer of the methyl group from (6S)-methyltetrahydrofolate to the corrinoid/iron–sulfur protein catalyzed by the methyltransferase from *Clostridium thermoaceticum*: a key step in the Wood–Ljungdahl pathway of acetyl-CoA synthesis. *Biochemistry* **38**, 5728–5735 (1999).
 24. Jarrett, J. T. *et al.* Mutations in the B₁₂-binding region of methionine synthase: how the protein controls methylcobalamin reactivity. *Biochemistry* **35**, 2464–2475 (1996).
 25. Bandarian, V., Ludwig, M. L. & Matthews, R. G. Factors modulating conformational equilibria in large modular proteins: a case study with cobalamin-dependent methionine synthase. *Proc. Natl. Acad. Sci. USA* **100**, 8156–8163 (2003).
 26. Wirt, M. D. *et al.* Structural and electronic factors in heterolytic cleavage: formation of the Co(I) intermediate in the corrinoid/iron–sulfur protein from *Clostridium thermoaceticum*. *Biochemistry* **34**, 5269–5273 (1995).
 27. Stich, T. A. *et al.* Spectroscopic studies of the corrinoid/iron–sulfur protein from *Moorella thermoacetica*. *J. Am. Chem. Soc.* **128**, 5010–5020 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. E. Darty for his assistance with the purification of CFeSP. This work was supported by National Institutes of Health grants GM69857 (to C.L.D.) and GM39451 (to S.W.R.) and the MIT Energy Initiative (to C.L.D.). C.L.D. is a Howard Hughes Medical Institute Investigator. This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the US Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Author Contributions Y.K. performed crystallization and data collection, processing and refinement that gave the folate-free and folate-bound CFeSP/MeTr structures. N.A. built the microspectrophotometer and performed *in crystallo* spectroscopic experiments with the aid of Y.K., who performed the parallel solution spectroscopic experiments. T.I.D. determined initial crystallization conditions and performed initial data collection, and L.C.B. processed and refined these data. G.B. and J.S. expressed and purified protein samples, and S.W.R. and C.L.D. were involved in study design. Y.K. and C.L.D. wrote the manuscript.

Author Information Atomic coordinates are deposited in the Protein Data Bank under accession codes 4DJD, 4DJF and 4DJF. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.L.D. (cdrennan@mit.edu).

METHODS

Protein purification. CFeSP was expressed and purified anaerobically from *M. thermoacetica* ATCC 39073 as described²¹, except for the following modifications. CFeSP was purified from cell extracts using DEAE-cellulose and high-resolution Q-Sepharose anion exchange chromatography followed by phenyl-Sepharose hydrophobic interaction chromatography. Fractions containing CFeSP were concentrated and buffer exchanged using Amicon ultracentrifuge concentrators in the anaerobic chamber. MeTr was expressed and purified anaerobically from recombinant *E. coli* as described⁶. Concentrations of CFeSP and MeTr protein samples were determined using the Rose-Bengal method²⁸ and kept in storage buffer: 50 mM Tris-HCl, pH 7.6, 100 mM NaCl, 2 mM dithiothreitol.

Crystallization. Crystals of the folate-free CFeSP/MeTr complex were grown by hanging-drop vapour diffusion in an anaerobic chamber (Coy Laboratories) at room temperature by adding 1 μ l of precipitant (100 mM Bis-Tris, pH 6.5, 100 mM calcium acetate, 9% PEG 5,000 monomethyl ether, 20% glycerol) to 2 μ l of an equimolar mixture of CFeSP and MeTr (approximately 250 μ M monomer for each), over a 0.5 ml reservoir solution of precipitant. Large, brown, rod-shaped crystals appeared overnight. Crystals were looped and cryo-cooled in liquid nitrogen anaerobically before collection of X-ray diffraction data at 100 K. Crystals of CFeSP/MeTr co-crystallized with the CH₃-H₄folate substrate were obtained in the same manner as above, except the protein solution also contained CH₃-H₄folate at 1 mM concentration. Crystals of CFeSP/MeTr co-crystallized with both CH₃-H₄folate and Ti(III) citrate as a reductant were obtained in the same manner, except the precipitant solution also contained Ti(III) citrate at 3 mM concentration.

Structure determination of folate-free CFeSP/MeTr structure. Two X-ray diffraction data sets were collected for the folate-free CFeSP/MeTr structure. A lower-resolution data set (3.3 \AA) was collected at the Advanced Light Source (ALS) beam line 5.0.2 ($\lambda = 1.1000 \text{\AA}$), and a higher-resolution data set (2.38 \AA) was later collected at the Advanced Photon Source (APS) beam line 24ID-C ($\lambda = 0.9792 \text{\AA}$).

The initial data set to 3.3 \AA resolution was processed in HKL2000 and Scalepack²⁹. The structure was solved by molecular replacement in Phaser³⁰, using individual structures of MeTr¹⁰ (Protein Data Bank accession number 2E7F) and ChCFeSP⁹ (Protein Data Bank accession number 2H9A) lacking its B₁₂ domain as independent search models. Two CFeSP/MeTr complexes (approximately 220 kDa each) were found in the asymmetric unit, and crystals belonged to the space group P2₁2₁2₁ with unit cell dimensions (\AA): $a = 137.42$, $b = 159.87$ and $c = 241.92$. Iterative rounds of refinement with residue-grouped B-factors were performed in CNS³¹ and PHENIX³², with model building in Coot³³. The four B₁₂ domains present in the asymmetric unit were kept as a polyalanine model. Final R-factors for working and test reflections (R_{work} and R_{free}) were 29.2% and 33.7%, respectively, when refinement of the structure to higher resolution began. Data collection and refinement statistics for this data set are shown in Supplementary Table 1. Ramachandran analysis was performed in PROCHECK³⁴: 74.0% of residues resided in the most favoured region, with 21.2% additionally allowed, 3.1% generously allowed and 1.7% disfavoured.

The data set to 2.38 \AA resolution was processed in HKL2000 and Scalepack²⁹. Although this crystal formed in similar conditions as the crystal which gave the 3.3 \AA resolution data set, the space group was now P2₁2₁2, with unit cell dimensions (\AA): $a = 125.71$, $b = 242.84$ and $c = 79.67$. The structure for this crystal was thus solved by molecular replacement in Phaser³⁰ using the MeTr homodimer and CFeSP heterodimers lacking B₁₂ domains from the previously refined model of the 3.3 \AA resolution structure as independent search models. Only one CFeSP/MeTr complex was present in the asymmetric unit. Iterative rounds of refinement were performed in CNS³¹ and PHENIX³², with model building in Coot³³. Translation/libration/screw refinement was performed in latter refinement rounds with seven translation/libration/screw groups: the MeTr homodimer (chains A and B), the Fe₄S₄ domain of one CFeSP large subunit (chain C), the TIM domain of chain C with one small subunit (chain D), the B₁₂ domain of chain C, the Fe₄S₄ domain of the second large subunit (chain E), the TIM domain of chain E with the second small subunit (chain F), and the B₁₂ domain of chain E. Data collection and refinement statistics are shown in Supplementary Table 1, and average B-factors for each domain of the final model are given in Supplementary Table 2. Ramachandran analysis was performed in PROCHECK³⁴: 90.1% of residues reside in the most favoured region, with 9.5% additionally allowed, 0.3% generously allowed and 0.2% disfavoured. The final model contains residues 1–262 (of 262) for both MeTr chains (A and B), residues 2–442 (of 446) for both CFeSP large subunit chains (C and E) and residues 1–323 (of 323) for both CFeSP small subunit chains (D and F).

Except for the Fe₄S₄ and B₁₂ domains, the entire structure is composed of TIM barrels for which the electron density is well-defined (Supplementary Fig. 16). Electron density is weaker for the Fe₄S₄ domains (Supplementary Figs 5, 17 and

19), consistent with the fact that these domains exhibit higher B-factors (Supplementary Table 2 and Supplementary Fig. 7). However, reasonable electron density is present for the main chain and most side chains, allowing us to build a model for this domain. Still, several side chains of the Fe₄S₄ domains lack clear electron density; thus, for these residues, atoms were truncated past the C β atom (chain C, 12 residues truncated and chain E, 16 residues truncated, out of 56 total residues).

Although B-factors are high and electron density is weak for the B₁₂ domains in general, electron density for the B₁₂ cofactors is unambiguous (Supplementary Fig. 20), and density is also clear in several helical regions, including those near B₁₂ (Supplementary Fig. 18). Because the structure of a CFeSP B₁₂ domain bound with B₁₂ was already known⁹, we used the clear electron density of the B₁₂ cofactor and the resolvable helices to position the B₁₂ domain during model building. Still, many side chains of the CFeSP B₁₂ domains lacked clear electron density, and thus for these residues, atoms were truncated past the C β atom (chain C, 50 residues truncated and chain E, 59 residues truncated, out of 118 total residues).

The B₁₂ cofactor in the final model contains 5,6-dimethylbenzimidazole as the lower ligand moiety, as in cobalamin. Although active with cobalamin³⁵, previous studies have shown that CFeSP isolated from *M. thermoacetica* harbours an unusual B₁₂ derivative that contains 5'-methoxybenzimidazole as the lower ligand instead³⁶. However, disorder of the B₁₂ cofactor and B₁₂ domain owing to thermal motion of these regions in the CFeSP/MeTr crystal resulted in weak electron density for substituents of the benzimidazolyl ring (Supplementary Figs 18 and 20). Therefore, we cannot confirm the presence of this unusual B₁₂ derivative from our crystallographic studies, and we have thus modelled cobalamin as the form of B₁₂ in the structure.

Previous spectroscopic studies²⁷ in addition to the crystal structure of ChCFeSP⁹ have indicated that a water molecule coordinates the central cobalt of B₁₂ in the as-isolated CFeSP. Here, Co(II) is the major species and is expected to be five-coordinate. However, because of disorder we do not observe electron density to suggest a water molecule bound to cobalt (Supplementary Figs 18 and 20). Accordingly, we have not modelled a water molecule.

Structure determination of folate-bound CFeSP/MeTr structures. For crystals grown with CH₃-H₄folate, X-ray diffraction data were collected at APS beam line 24ID-C to 3.50 \AA resolution at $\lambda = 1.6039 \text{\AA}$ to optimize the cobalt peak anomalous signal. For crystals grown with both CH₃-H₄folate and Ti(III) citrate, X-ray diffraction data were collected at ALS beam line 8.2.2 to 3.03 \AA resolution at $\lambda = 1.0000 \text{\AA}$. The structures were solved by molecular replacement using the MeTr homodimer and CFeSP heterodimers lacking Fe₄S₄ and B₁₂ domains from the folate-free 2.38 \AA CFeSP/MeTr structure as independent search models. Refinement of the folate-free CFeSP/MeTr structure against either folate-bound X-ray data set was not sufficient to solve the structure, as the unit cell dimensions were markedly different (Supplementary Table 1). After molecular replacement, one CFeSP/MeTr complex was present in the asymmetric unit, and omit electron density clearly indicated the presence of bound folate (Supplementary Fig. 21). Iterative rounds of refinement were performed in CNS³¹ and PHENIX³², with model building in Coot³³. The same test set of reflections for R_{free} calculations was used for both folate-bound data sets. Data collection and refinement statistics are shown in Supplementary Table 1. Ramachandran analysis was performed in PROCHECK³⁴: for the CH₃-H₄folate-only structure, 89.6% of residues reside in the most favoured region, with 9.8% additionally allowed, 0.3% generously allowed and 0.2% disfavoured. For the CH₃-H₄folate with Ti(III) citrate structure, 89.5% of residues reside in the most favoured region, with 10.0% additionally allowed, 0.2% generously allowed and 0.3% disfavoured. The final models both contain folate, B₁₂ and residues 1–262 (of 262) for MeTr chains (A and B), residues 2–442 (of 446) for CFeSP large subunit chains (C and E) and residues 1–323 (of 323) for CFeSP small subunit chains (D and F). As with the folate-free structure, several side chains of the Fe₄S₄ domains for both folate-bound structures lacked clear electron density; thus for these residues, atoms were truncated past the C β atom (chain C, 15 residues truncated and chain E, 18 residues truncated, out of 52 total residues). Similarly, many side chains of the B₁₂ domains lacked electron density, and thus for these residues, atoms were truncated past the C β atom (chain C, 51 residues truncated and chain E, 78 residues truncated, out of 118 total residues). The liganded/oxidation states of folate and B₁₂ in these structures were determined by use of a microspectrophotometer (see below).

Solution and *in crystallo* ultraviolet-visible absorption spectroscopy to determine enzyme activity *in crystallo*. Titanium(III) citrate (100 mM in 50 mM Tris, pH 7.6) was prepared³⁷, and (6S)-5-methyl-5,6,7,8-tetrahydrofolate (CH₃-H₄folate) containing one glutamate tail was purchased from Schircks Laboratories. As-isolated, reduced and methylated CFeSP samples in solution were prepared in a room-temperature anaerobic chamber (MBraun) following similar procedures to those previously described^{18,19,22,23,26,27}. Briefly, purified CFeSP (20 μ M) was used for the as-isolated sample, CFeSP mixed with Ti(III) citrate (1 mM) was used for

the reduced sample, and CFeSP mixed with equimolar MeTr, Ti(III) citrate (1 mM), and CH₃-H₄folate (1 mM) was used for the methylated sample. Spectra were taken using a Nanodrop 2000c (Thermo Scientific) in a quartz cuvette or on the sample stage in the anaerobic chamber directly after mixing; identical solutions lacking CFeSP were used as blanks.

To obtain *in crystallo* absorption spectra, CFeSP/MeTr crystals in as-isolated, reduced and methylated forms were prepared in a similar fashion. In a room-temperature anaerobic chamber (Coy Laboratories), crystals were looped into a 2 μ l drop, which was placed on a cover slide and contained one of the following three solutions for as-isolated, reduced and methylated samples, respectively: well solution, well solution with Ti(III) citrate (10 mM) and well solution with Ti(III) citrate (10 mM) and CH₃-H₄folate (1 mM). A ring of epoxy surrounding each drop was applied to the cover side, and a second cover slide was placed on top, sandwiching the drops within a uniform distance separation and sealing the crystals within an anaerobic environment. Upon curing of the epoxy, crystals were brought out of the anaerobic chamber and mounted on an XZ translation stage (Newport, UMR8.25 and SM-13) in a fibre optic coupled microspectrophotometer (Ocean Optics, Jaz) with 40 mm diameter reflective objectives (Optique Peter) and a deuterium-halogen lamp (DH2000-BAL, Ocean Optics), similar to that previously described^{38,39}. Stray light was blocked with blackout material. The light focus was coarsely aligned to the crystals by visual inspection and then finely aligned by monitoring light transmission in real time. Data were acquired at room temperature with the SpectraSuite software (Ocean Optics). The background transmission was measured through the solution immediately surrounding the crystals. The dark current was measured with the light shuttered off. Sample, reference and dark current spectra were acquired by averaging 10–50 scans with total exposure times of 90–1000 ms. Experiments were completed within 60 min of sample preparation, and crystals remained intact over the course of the experiment, as observed using a microscope after data collection.

To generate Fig. 3, absorbance spectra were scaled relative to each other to account for variable crystal sizes and path lengths, where absolute peak absorbances did not exceed one absorbance unit.

In crystallo ultraviolet-visible absorption spectroscopy on folate-bound CFeSP/MeTr crystals to determine liganded/oxidation state of bound B₁₂ and folate. Ultraviolet-visible absorption spectra were collected on a microspectrophotometer at 100 K for crystals of folate-free CFeSP/MeTr, crystals that were grown in the presence of CH₃-H₄folate only and for crystals grown in the presence of both CH₃-H₄folate and Ti(III) citrate (Supplementary Fig. 9). The spectra were compared with the analogous solution spectra (Fig. 3). The spectrum for the folate-free crystal was similar to the spectrum of CFeSP alone in solution,

with broad features at approximately 400 and 470 nm indicative of the Fe₄S₄ cluster and B₁₂ primarily in the Co(II) state. The spectrum for the crystal grown with CH₃-H₄folate matched the spectrum of folate-free crystals, indicating that B₁₂ had remained primarily in the Co(II) state, and turnover had not occurred. However, the spectrum for the crystal grown with both CH₃-H₄folate and Ti(III) citrate was markedly different and contained a peak at 450 nm, indicating that B₁₂ in these crystals was methylated to the CH₃-Co(III) state. Based on these data, we modelled the methyl group on folate in the structure co-crystallized with CH₃-H₄folate only, whereas we modelled the methyl group bound to Co of B₁₂ for the CH₃-H₄folate/Ti(III) citrate structure. Without these spectroscopic data, assignment of the location of the methyl group would otherwise have been prevented by the resolution limits of the data.

28. Elliott, J. I. & Brewer, J. M. The inactivation of yeast enolase by 2,3-butanedione. *Arch. Biochem. Biophys.* **190**, 351–357 (1978).
29. Otwowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
30. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
31. Brünger, A. T. *et al.* *Crystallography & NMR System*: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
32. Adams, P. D. *et al.* PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
33. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
34. Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
35. Lu, W.-P., Schiau, I., Cunningham, J. R. & Ragsdale, S. W. Sequence and expression of the gene encoding the corrinoid/iron–sulfur protein from *Clostridium thermoaceticum* and reconstitution of the recombinant protein to full activity. *J. Biol. Chem.* **268**, 5605–5614 (1993).
36. Ljungdahl, L. G., LeGall, J. & Lee, J.-P. Isolation of a protein containing tightly bound 5-methoxybenzimidazolylcobamide (factor IIIm) from *Clostridium thermoaceticum*. *Biochemistry* **12**, 1802–1808 (1973).
37. Zehnder, A. J. B. & Wuhrmann, K. Titanium(III) citrate as a nontoxic oxidation-reduction buffering system for the culture of obligate anaerobes. *Science* **194**, 1165–1166 (1976).
38. Royant, A. *et al.* Advances in spectroscopic methods for biological crystals. 1. fluorescence lifetime measurements. *J. Appl. Cryst.* **40**, 1105–1112 (2007).
39. Barstow, B., Ando, N., Kim, C. U. & Gruner, S. M. Alteration of citrine structure by hydrostatic pressure explains the accompanying spectral shift. *Proc. Natl. Acad. Sci. USA* **105**, 13362–13366 (2008).

Structure and mechanism of a glutamate–GABA antiporter

Dan Ma^{1,3}, Peilong Lu^{1,3}, Chuangye Yan^{2,3}, Chao Fan¹, Ping Yin², Jiawei Wang² & Yigong Shi^{1,3}

Food-borne hemorrhagic *Escherichia coli*, exemplified by the strains O157:H7 and O104:H4 (refs 1, 2), require elaborate acid-resistance systems (ARs)³ to survive the extremely acidic environment such as the stomach (pH ≈ 2). AR2 expels intracellular protons through the decarboxylation of L-glutamate (Glu) in the cytoplasm and exchange of the reaction product γ-aminobutyric acid (GABA) with extracellular Glu. The latter process is mediated by the Glu–GABA antiporter GadC^{4,5}, a representative member of the amino-acid–polyamine–organocation superfamily of membrane transporters. The functional mechanism of GadC remains largely unknown. Here we show, with the use of an *in vitro* proteoliposome-based assay, that GadC transports GABA/Glu only under acidic conditions, with no detectable activity at pH values higher than 6.5. We determined the crystal structure of *E. coli* GadC at 3.1 Å resolution under basic conditions. GadC, comprising 12 transmembrane segments (TMs), exists in a closed state, with its carboxy-terminal domain serving as a plug to block an otherwise inward-open conformation. Structural and biochemical analyses reveal the essential transport residues, identify the transport path and suggest a conserved transport mechanism involving the rigid-body rotation of a helical bundle for GadC and other amino acid antiporters.

Other homologous amino-acid–polyamine–organocation (APC) family members include the key AR3 component arginine–agmatine (Arg–Agm) antiporter AdiC^{6,7} (Fig. 1a), the lysine–cadaverine antiporter CadB, and the putrescine–ornithine antiporter PotE. Structural analysis of AdiC^{8–11} revealed a conserved LeuT fold that is associated with the Na⁺-coupled symporters^{12–15} and a proton-coupled transporter ApcT¹⁶. Ligand-free AdiC exists in an outward-open conformation^{8,10}, and binding of Arg triggers a major structural rearrangement resulting in an occluded conformation⁹. Despite these advances, the transport mechanism for AdiC or any other amino acid antiporter remains poorly understood. The inward-open conformation has yet to be captured for any amino acid antiporter.

The full-length, wild-type (WT) GadC (residues 1–511), derived from the *E. coli* strain O157:H7, was purified to homogeneity. To characterize GadC, we reconstituted a proteoliposome-based transport assay (Fig. 1b), in which substrate transport was monitored through the detection of ³H-labelled Glu. Incorporation of GadC into the liposome allowed the rapid accumulation of Glu at pH 4.5 (Fig. 1b; Supplementary Fig. 1). Strikingly, substrate transport by GadC decreased sharply with increasing pH values; the accumulation of Glu after 15 s at pH 5.0, 5.5 and 6.0 was about 67%, 32% and 8% of that at pH 4.5 (Supplementary Fig. 1). At pH 6.5 and above, there was no detectable accumulation of Glu (Fig. 1b). Thus, substrate transport by GadC is strictly pH-dependent, with robust activity at pH 5.5 or below. This biochemical property not only helps in the survival of enterobacteria under acidic environment, but it may also be important for avoiding unnecessary proton efflux under neutral pH growth conditions.

AdiC has been thoroughly characterized, with elevated transport activities at pH 4 and 6 (ref. 17); however, AdiC still allows moderate

transport even under basic conditions such as pH 8 (ref. 17). We compared the activities of AdiC and GadC under identical sets of experimental conditions (Fig. 1c, d). The results confirmed and extended published findings. The total substrate accumulation in the proteoliposomes over 10 min at pH 9.0 was about 21% of that at pH 5.0 (Fig. 1c). In sharp contrast, GadC has no detectable activity at pH 6.5 or higher. Determination of the maximal transport activity (V_{max}) and K_m for AdiC and GadC at different pH values confirmed these observations (Fig. 1d and Supplementary Fig. 2). Thus, in comparison with AdiC, GadC exhibits much more stringent dependence on pH for substrate transport.

Next we replaced GABA individually with 19 natural amino acids; the substrate transport for each amino acid was measured at pH 5.5 (Fig. 1e). In addition to Glu, GadC also efficiently transports three additional amino acids: Gln and, to a smaller extent, Met and Leu. This result was unexpected because, unlike Glu, none of the three amino acids carries a charge or can be protonated on the side chain. Notably, the V_{max} for Gln is considerably larger than that for Glu or GABA at pH 5.5, whereas the K_m for Gln is comparable to that for GABA (Supplementary Fig. 3). In particular, GadC allows little, if any, transport of the amino acids Asp, Phe, Gly, His, Lys, Asn, Pro and Trp (Fig. 1e). Substrate transport for Asp and Asn was less than 5% of that for the chemically similar amino acids Glu and Gln, respectively. This analysis strongly indicates that GadC is highly selective in substrate transport.

Efflux of protons presumably contributes to a build-up of a negative potential within the lipid membrane, which is known to be unfavourable for continued substrate transport by amino acid antiporters such as AdiC¹⁷ and ApcT¹⁸. We examined Glu–GABA exchange in the proteoliposome-based assay in the absence or presence of valinomycin (Supplementary Fig. 4). The results clearly demonstrate that positive potential within the proteoliposomes stimulated substrate transport, whereas negative potential led to decreased transport.

We crystallized the full-length GadC in the space group $P2_12_12_1$ at pH 8.0 and determined the structure by platinum-based single-wavelength anomalous dispersion (SAD) at 3.1 Å resolution (Supplementary Tables 1 and 2; Supplementary Figs 5 and 6). Each asymmetric unit contains two molecules of GadC, arranged in an antiparallel fashion (Supplementary Fig. 5). This packing pattern suggests that, similarly to AdiC¹⁰ and the transporter CIC¹⁹, the functional entity for GadC is likely to be a single molecule. GadC contains 12 TMs, with TM1 and TM6 each containing two short α -helices connected by a discontinuous stretch in the middle (Fig. 2a and Supplementary Fig. 7). The structure of GadC, together with the identification of two periplasmic loops in GadC²⁰, allows an unambiguous assignment of its membrane topology. In a similar manner to AdiC and other LeuT-type transporters, GadC contains two inverted repeats, TM1–TM5 and TM6–TM10, which are related to each other by a pseudo-two-fold axis (Supplementary Fig. 8a). The periplasmic and cytoplasmic sides of GadC are highly charged (Supplementary Fig. 8b).

¹Ministry of Education Protein Science Laboratory, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. ²State Key Laboratory of Biomembrane and Membrane Biotechnology, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. ³Tsinghua–Peking Joint Center for Life Sciences, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China.

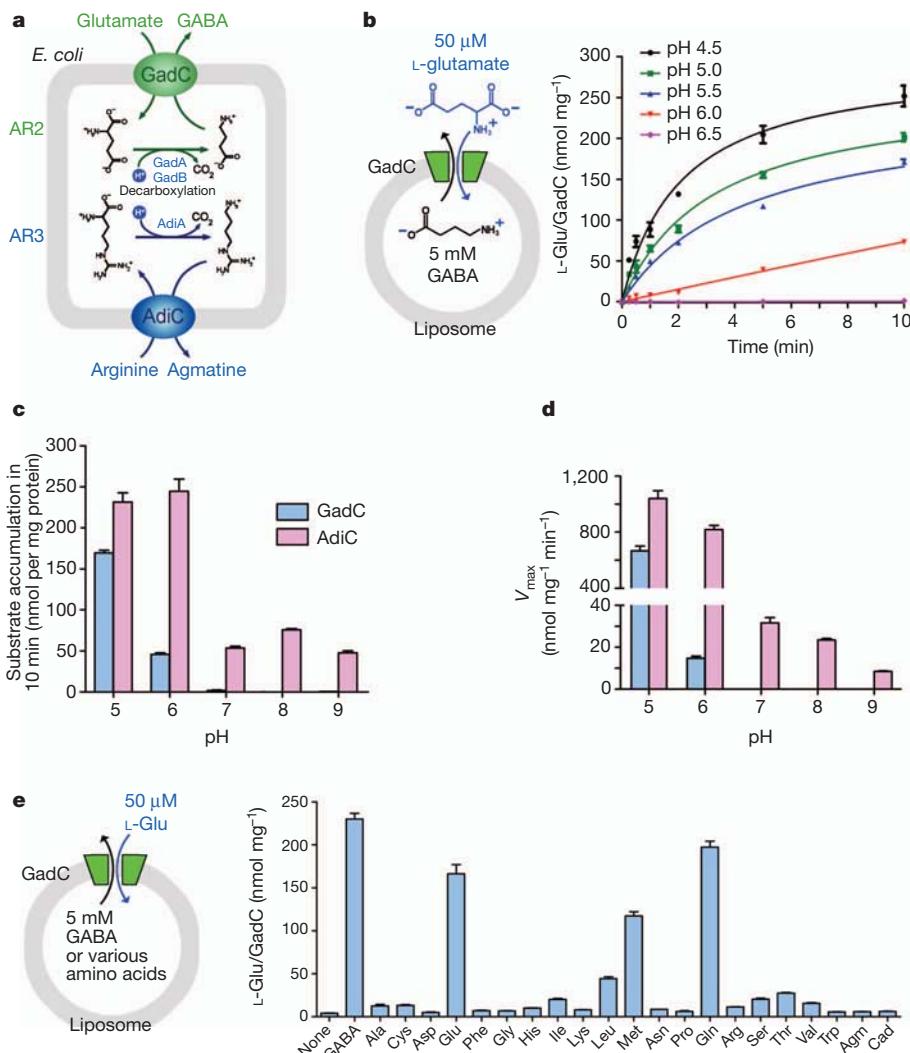


Figure 1 | Functional characterization of GadC. **a**, Schematic diagram of AR2 and AR3 in *E. coli*. The antiporter GadC of AR2 exchanges extracellular Glu for intracellular GABA, resulting in the net efflux of one proton per cycle. Glu is decarboxylated by GadA/B to become GABA in cells. The Arg–Agn antiporter AdiC and the decarboxylase AdiA are the equivalents of GadC and GadA/B, respectively. **b**, GadC shows pH-dependent transport in the proteoliposome assay. ^3H -labelled Glu was present at roughly $0.19\text{ }\mu\text{M}$ to permit the measurement of transport. The transport activity is robust at pH 4.5 and rapidly decreases with increasing pH, with no detectable transport activity at pH 6.5 or above. **c**, Comparison of pH-dependent substrate transport of AdiC and GadC. GadC exhibits stringent pH dependence, with no detectable transport activity at pH values higher than 7.0. By contrast, AdiC has considerable transport activity at pH 7.0, 8.0 and 9.0. **d**, Comparison of V_{max} for AdiC and GadC at different pH values. **e**, The substrate specificity of GadC-mediated transport in the proteoliposome assay. Substrate transport was measured at pH 5.5 for 10 min. GadC only allows transport of GABA, Glu, Gln and (to a smaller extent) Met and Leu. All error bars represent the s.d. for three independent experiments.

In comparison with AdiC and other homologues, GadC contains a unique, extended C-terminal fragment (Supplementary Fig. 9). In contrast with reported structures of AdiC^{8–11}, GadC seems to adopt an inward-open conformation (Fig. 2a). The open path leads to a negatively charged environment (Fig. 2b), where substrate-binding residues are likely to be located. However, the C-terminal fragment (residues 477–511) forms a folded domain and completely blocks the path to the putative substrate-binding site. The C-terminal fragment, with clear electron density (Supplementary Fig. 10a), is hereafter referred to as the C-plug. The observation of a blocked transport path in GadC is consistent with the fact that the crystals were generated at pH 8.0, at which no transport activity could be detected (Fig. 1c).

The location of the C-plug strongly suggests that its displacement is a prerequisite for the transport activity of GadC. To examine this, we generated a C-plug-deleted GadC variant (residues 1–470) and measured its ability to permeate Glu–GABA. Strikingly, whereas WT GadC showed little substrate transport over 60 min at pH 6.5, GadC (1–470) showed a significant level of Glu accumulation (Fig. 3a).

Additional measurements show that the transport activity of GadC (1–470) was rapidly decreased with increasing pH and became undetectable at pH 7.5 or above (Fig. 3b). Overall, deletion of the C-plug in GadC shifted its pH-dependent substrate transport towards a higher pH.

The C-plug contains several basic amino acids and makes multiple intra-domain and inter-domain hydrogen bonds (Fig. 3c and Supplementary Fig. 10b). The tightly folded conformation of the C-plug is stabilized by two centrally located basic residues, His 491 and Arg 499. His 491 donates a hydrogen bond to Ser 500, whereas Arg 499 makes three hydrogen bonds to the main-chain atoms of Phe 492 and Leu 494. At one end of the plug, Tyr 503 forms a hydrogen bond to Ala 487. These intra-domain interactions are complemented by inter-domain contacts. At the other end of the plug, the guanidinium group of Arg 497 donates a hydrogen bond to Gln 98 on TM3. In addition, His 502 makes a hydrogen bond to Arg 314 on TM8, whereas the main chain atoms of Val 477 and Ser 484 form hydrogen bonds with Gln 321 on TM8 and Glu 226 on TM6.

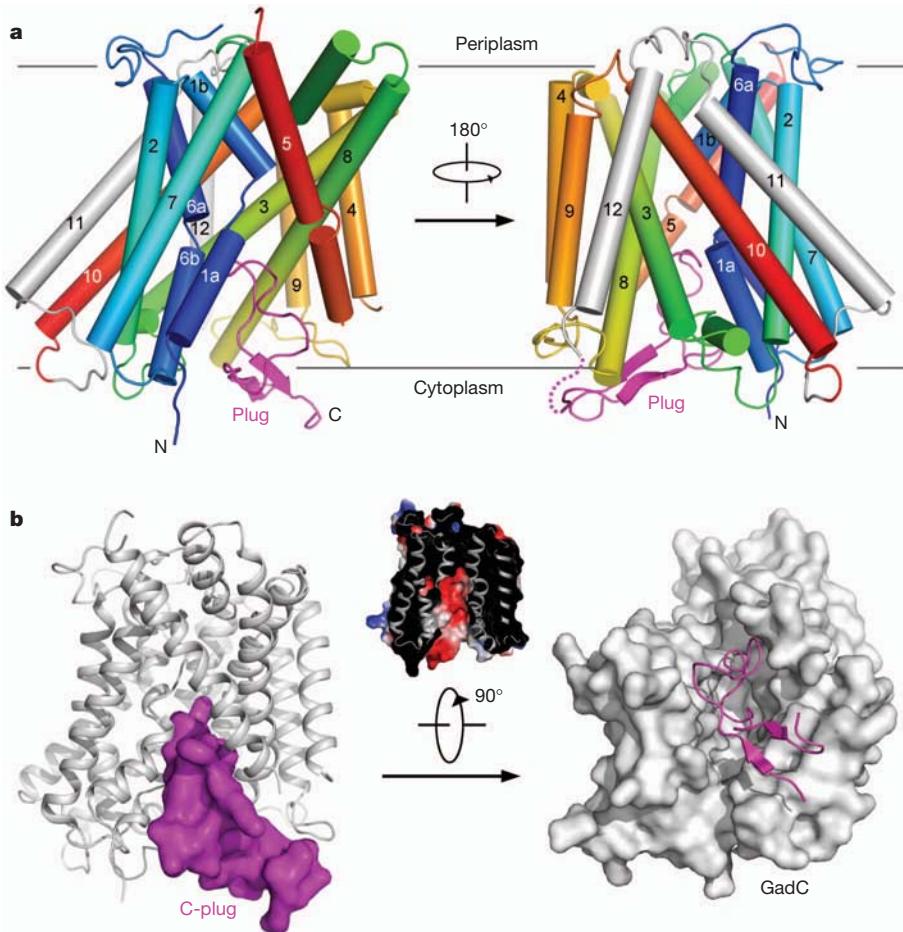


Figure 2 | Overall structure of GadC. **a**, Overall structure of the WT full-length GadC. TM1–TM10 are rainbow-coloured, with TM1 in blue and TM10 in red. TM11 and TM12 are shown in grey. The C-terminal fragment (C-plug) is coloured magenta. **b**, The C-plug of GadC blocks an otherwise inward-open

conformation. Two perpendicular views of GadC are shown. The C-plug blocks access to the negatively charged substrate-binding cleft (inset). All structural figures were prepared with PyMol³⁰.

Perturbation of these interactions is predicted to compromise the stability of the C-plug and consequently to facilitate its displacement. To examine this prediction, we generated five GadC variants—H491A,

R497A, R499A, H502A and Y503A—each carrying a missense mutation on a critical residue in the C-plug. We individually investigated substrate transport by these GadC variants at pH 6.5. Whereas the WT GadC exhibited little activity, all five GadC variants allowed varying but significant levels of substrate transport (Fig. 3d). In particular, the transport activities for GadC-H491A and GadC-R497A were similar to that of GadC (1–470) (Fig. 3d). By contrast, two additional GadC variants, H495A and H511A, allowed considerably less Glu accumulation, consistent with the observation that these mutations do not affect any critical interactions in the C-plug.

The closed conformation of GadC is attained not only by the C-plug in the cytoplasm but also by the L7 loop at the periplasmic side (Supplementary Fig. 11a). The L7 loop interacts with surrounding structural elements through a combination of hydrogen bonds and van der Waals contacts (Supplementary Fig. 11b, c). We speculate that,

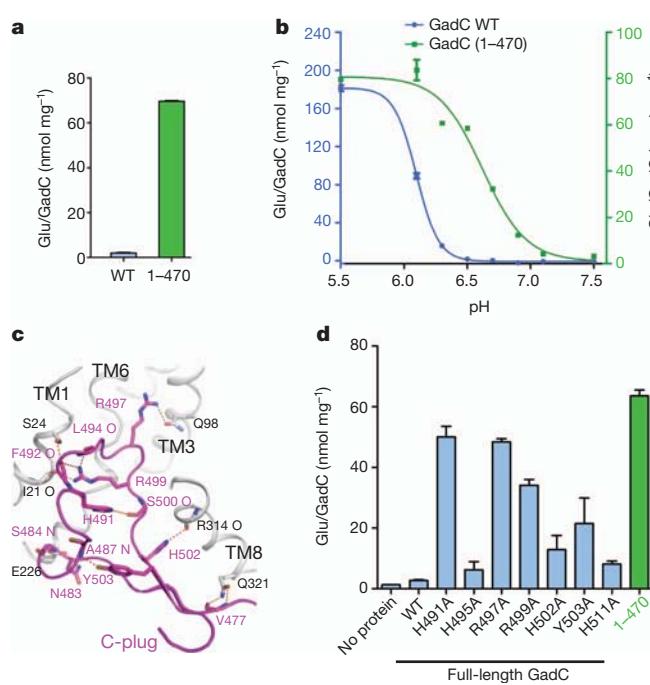


Figure 3 | The C-plug regulates substrate transport. **a**, Truncation of the C-plug (residues 471–511) allowed transport of Glu at pH 6.5. Shown here is the total accumulation of Glu in the proteoliposomes in 60 min. **b**, Deletion of the C-plug shifts the pH-dependent transport activity of GadC towards higher pH values. **c**, The C-plug interacts with surrounding structural elements through multiple hydrogen bonds. The C-plug is coloured magenta and surrounding structural elements are shown in grey. **d**, Structural integrity of the C-plug is important for the transport activity of the WT GadC. Although the WT GadC does not allow apparent transport of substrates at pH 6.5, several missense mutants acquired this ability. These mutations probably led to compromised interactions between the C-plug and its surrounding structural elements. All error bars represent the s.d. for three independent experiments.

during each cycle of transport, the L7 loop must be displaced, at least transiently, to allow the passage of substrate molecules. The $\text{C}\alpha$ - $\text{C}\alpha$ distance between residue 267 on L7 and residue 364 on TM10 is about 5.6 Å (Supplementary Fig. 11d), which is ideal for disulphide bond formation if these two residues are replaced by Cys. We generated a double mutation GadC-L267C/N364C and subjected the purified protein to oxidation by *o*-phenanthroline copper complex²¹. The oxidized GadC variant (L267C/N364C) showed undetectable substrate transport in the absence of the reducing agent dithiothreitol but restored substrate transport in the presence of dithiothreitol (Supplementary Fig. 11d). By contrast, the oxidized WT GadC showed similar levels of substrate transport in the absence or presence of dithiothreitol.

The available structures of AdiC⁸⁻¹¹ greatly facilitated the identification of gating residues in GadC. The substrate transport path is sandwiched axially between the C-plug and the L7 loop and surrounded laterally by TM1, TM3, TM6, TM8 and TM10 (Fig. 4a). Sequence alignment with AdiC and other amino acid antiporters identified six potential gating residues in GadC: Tyr 96, Tyr 214, Glu 218, Trp 308, Tyr 378 and Tyr 382 (Supplementary Fig. 9). These residues are located within or in close proximity to the putative transport path (Fig. 4a). Comparison of the spatial locations for these amino acids revealed a pattern that differed from that in AdiC of the outward-open conformation⁸ or Arg-bound occluded conformation⁹ (Supplementary Fig. 12).

The distal gate in AdiC comprises three amino acids, with Glu 208 hydrogen bonded to Tyr 93 and Tyr 365 (ref. 9). Glu 208 and Tyr 365 in AdiC correspond to Glu 218 and Tyr 382 in GadC, respectively (Supplementary Fig. 9). Tyr 378 is also located within hydrogen-bonding distance of Glu 218. Tyr 96 in GadC, which may correspond to Tyr 93 in AdiC, is located roughly 5 Å away from the position of Tyr 93

(Supplementary Fig. 12). The middle gate residue Trp 293 in AdiC corresponds to Trp 308 in GadC, which is farther from the axial centre of transport, probably reflecting the nature of the inward-open conformation. The proximal gate residue, Trp 202, in AdiC closes on binding to Arg and occludes the substrate molecule from the periplasm⁹; the corresponding residue remains to be conclusively identified in GadC. The nearby residues Phe 210 and Tyr 214 are located in different positions from that of Trp 202 in AdiC. This analysis suggests major conformational changes for these putative gating residues during the substrate transport cycle. Several other amino acids are located within the putative transport path and may have a function in transport. In particular, Tyr 30 on TM1 donates a hydrogen bond to Glu 304 on TM8 (Supplementary Fig. 13a).

To corroborate the structural analysis, we generated six GadC variants—Y30A, E218A, E304A, W308A, Y378A and Y382A—each with a missense mutation targeting a putative gating residue. In contrast with WT GadC, each mutation caused at least 90% decrease in substrate transport (Supplementary Fig. 13b). GadC-W308A, which affects the putative middle gate residue Trp 308, only retained about 2% of the WT activity. By contrast, the GadC variant M25A, which does not affect any putative transport residue, showed about 74% of the WT transport activity.

Under the assumption of a conserved transport mechanism between GadC and AdiC, the inward-open conformation of GadC, with omission of the C-plug, probably reflects a distinct state of AdiC during substrate transport. Structural alignment between outward-open AdiC⁸ and GadC reveals a striking pattern of conformational changes that are concentrated in a helical bundle comprising TM1, TM2, TM6 and TM7 (Fig. 4b). For simplicity of discussion, these four TMs are collectively named the gate domain, and the rest is referred to

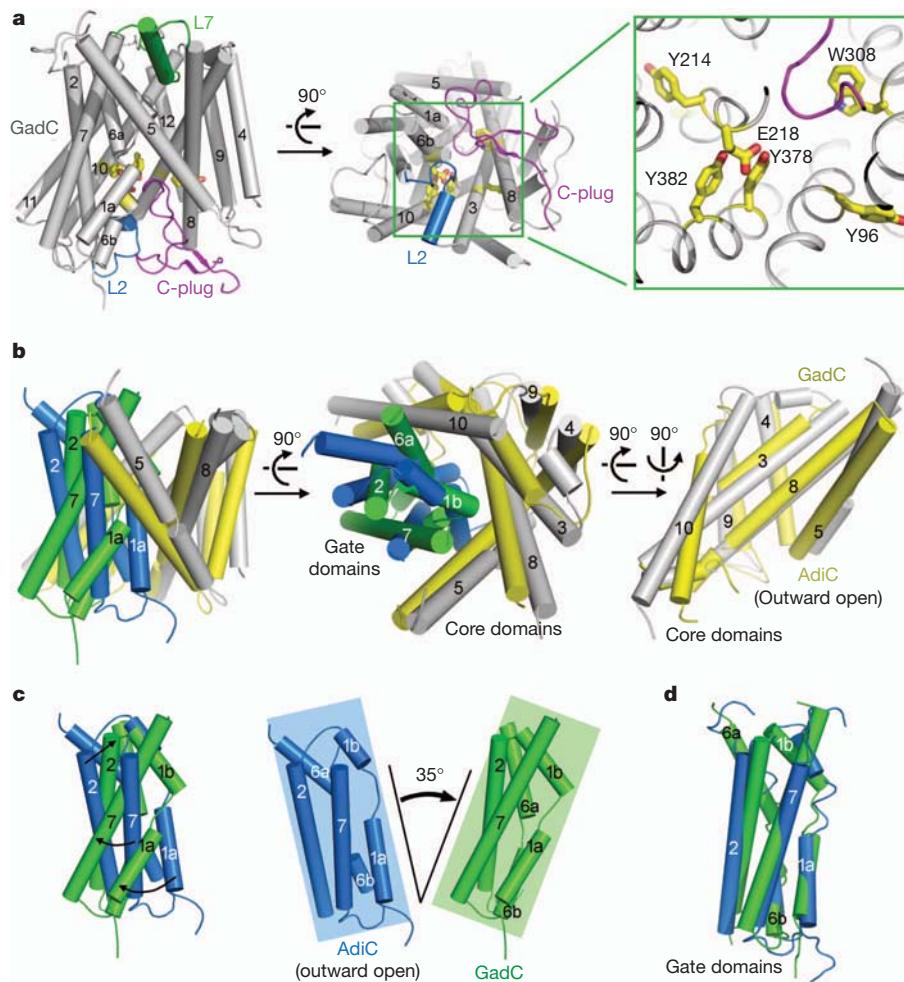


Figure 4 | The transport path in GadC.

a, Identification of key amino acids in the transport path. Sequence alignment and structural analysis identify six amino acids that may be essential in substrate transport. These residues are shown in a close-up view in the right panel. **b**, The gate domain in GadC, comprising TM1, TM2, TM6 and TM7, undergoes the most pronounced conformational changes compared with those of the outward-open AdiC⁸. The core domain showed only minor changes. The gate and core domains of GadC are coloured green and grey, respectively; the gate and core domains of AdiC are coloured blue and yellow, respectively. **c**, The conformational changes of the gate domain amount to a rigid-body rotation of about 35°. **d**, Structural overlay of the gate domains from GadC (green) and AdiC (blue).

as the core domain. In contrast with that in AdiC, the gate domain in GadC seems to rotate clockwise by about 35° (Fig. 4c), resulting in an outward-closed and inward-open conformation in GadC. Structural alignment between the isolated gate domains of AdiC and GadC revealed only minor changes in these four TMs (Fig. 4d), suggesting a rigid-body movement between the gate and the core domains. This structural analysis supports the notion of alternating access for membrane transporters^{22,23}. The four TMs in the gate domain undergo the most drastic structural rearrangement among all 12 TMs of GadC (Supplementary Fig. 14). In comparison with the gate domain, TMs of the core domain have a considerably smaller degree of conformational changes, particularly TM3, TM8 and TM9.

The mechanistic conservation may go beyond the amino acid antiporters. Mhp1 was recently found to switch from outward-open to inward-open conformation through a rigid-body rotation involving two helical bundles²⁴. After the exit of Na^+ , the Na^+ -galactose transporter vSGLT was also thought to undergo a minor rigid-body movement involving two helical bundles²⁵. Of the two helical bundles, one contains TM3, TM4, TM8 and TM9, and the other comprises TM2, TM6 and TM7; this is true for Mhp1 (ref. 24), vSGLT²⁵ and AdiC/GadC, for all of which at least two distinct conformations of a transport cycle have been structurally characterized. The helical bundle that executes the rigid-body movement is proposed to be the gate domain (TM1, TM2, TM6 and TM7) in GadC and the hash motif (TM3, TM4, TM8 and TM9) in Mhp1 (ref. 24). Thus, the moving and non-moving portions of GadC are exactly the reciprocal of those in Mhp1. The choice of the core domain in GadC is justified by the following analysis. Superposition between AdiC and GadC yields a root mean squared deviation of 2.98 Å over 167 aligned $\text{C}\alpha$ atoms if TM3, TM4, TM5, TM8, TM9 and TM10 are treated as the non-moving portion, and about 3.95 Å over 143 aligned $\text{C}\alpha$ atoms if TM1, TM2, TM5, TM6, TM7 and TM10 are treated as the non-moving portion.

Structural elucidation of the Glu–GABA antiporter GadC is a step towards a detailed, mechanistic understanding of the amino acid antiporters. Many questions remain (see Supplementary Discussion). At present we have little information about how and to what extent the C-plug of GadC is dislodged during substrate transport. We have yet to explain the pH-dependent transport activity by amino acid antiporters such as GadC and AdiC. Conclusive answers to these questions require additional biochemical and structural investigation.

METHODS SUMMARY

All constructs were generated by standard PCR-based protocol. GadC was over-expressed in *E. coli* BL21(DE3). The proteins were purified to homogeneity by affinity chromatography and gel filtration. Crystals of GadC were grown by the hanging-drop vapour-diffusion method. The crystals belong to the space group $P2_12_12_1$; they were flash-frozen in a cold nitrogen stream at 100 K. X-ray data were collected at the Shanghai Synchrotron Radiation Facility (SSRF) beamline BL17U and SPRing-8 beamline BL41XU. Data were processed with HKL-2000 (ref. 26). The platinum positions were determined with the program SHELXD²⁷. Cross-crystal averaging with all three data sets combined with solvent flattening, histogram matching and non-crystallographic symmetry (NCS) averaging in DMMulti gave a map of sufficient quality for model building. An initial model was built into the experimental map by using COOT²⁸. The sequence docking was aided with the selenium sites in the anomalous difference Fourier map. The structure was refined with PHENIX²⁹. Proteoliposome assays were performed to determine transport activities of various GadC mutants.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 October 2011; accepted 3 February 2012.

Published online 11 March 2012.

1. Riley, L. W. *et al.* Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308**, 681–685 (1983).
2. Rohde, H. *et al.* Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N. Engl. J. Med.* **365**, 718–724 (2011).
3. Foster, J. W. *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nature Rev. Microbiol.* **2**, 898–907 (2004).

4. Hersh, B. M., Farooq, F. T., Barstad, D. N., Blankenhorn, D. L. & Slonczewski, J. L. A glutamate-dependent acid resistance gene in *Escherichia coli*. *J. Bacteriol.* **178**, 3978–3981 (1996).
5. Castanie-Cornet, M. P., Penfound, T. A., Smith, D., Elliott, J. F. & Foster, J. W. Control of acid resistance in *Escherichia coli*. *J. Bacteriol.* **181**, 3525–3535 (1999).
6. Iyer, R., Williams, C. & Miller, C. Arginine–agmatine antiporter in extreme acid resistance in *Escherichia coli*. *J. Bacteriol.* **185**, 6556–6561 (2003).
7. Gong, S., Richard, H. & Foster, J. W. YjdE (AdiC) is the arginine:agmatine antiporter essential for arginine-dependent acid resistance in *Escherichia coli*. *J. Bacteriol.* **185**, 4402–4409 (2003).
8. Gao, X. *et al.* Structure and mechanism of an amino acid antiporter. *Science* **324**, 1565–1568 (2009).
9. Gao, X. *et al.* Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* **463**, 828–832 (2010).
10. Fang, Y. *et al.* Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* **460**, 1040–1043 (2009).
11. Kowalczyk, L. *et al.* Molecular basis of substrate-induced permeation by an amino acid antiporter. *Proc. Natl. Acad. Sci. USA* **108**, 3935–3940 (2011).
12. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^- -dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
13. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na^+ /betaaine symporter BetP. *Nature* **458**, 47–52 (2009).
14. Faham, S. *et al.* The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+ /sugar symport. *Science* **321**, 810–814 (2008).
15. Weyand, S. *et al.* Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709–713 (2008).
16. Shafer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. Structure and mechanism of a Na^+ -independent amino acid transporter. *Science* **325**, 1010–1014 (2009).
17. Fang, Y., Kolmakova-Partensky, L. & Miller, C. A bacterial arginine–agmatine exchange transporter involved in extreme acid resistance. *J. Biol. Chem.* **282**, 176–182 (2007).
18. Sasahara, A., Nanatani, K., Enomoto, M., Kuwahara, S. & Abe, K. Substrate specificity of the aspartate:alanine antiporter (AspT) of *Tetragenococcus halophilus* in reconstituted liposomes. *J. Biol. Chem.* **286**, 29044–29052 (2011).
19. Robertson, J. L., Kolmakova-Partensky, L. & Miller, C. Design, function and structure of a monomeric CIC transporter. *Nature* **468**, 844–847 (2010).
20. Waterman, S. R. & Small, P. L. Identification of sigma S-dependent genes associated with the stationary-phase acid-resistance phenotype of *Shigella flexneri*. *Mol. Microbiol.* **21**, 925–940 (1996).
21. Kobashi, K. Catalytic oxidation of sulphydryl groups by o-phenanthroline copper complex. *Biochim. Biophys. Acta* **158**, 239–245 (1968).
22. Abramson, J. *et al.* The lactose permease of *Escherichia coli*: overall structure, the sugar-binding site and the alternating access model for transport. *FEBS Lett.* **555**, 96–101 (2003).
23. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
24. Shimamura, T. *et al.* Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470–473 (2010).
25. Watanabe, A. *et al.* The mechanism of sodium and substrate release from the binding pocket of vSGLT. *Nature* **468**, 988–991 (2010).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
27. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
28. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
29. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
30. DeLano, W. L. The PyMOL Molecular Graphics System, (<http://www.pymol.org>) (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Huang and J. He (at SSRF beamline BL17U) and N. Shimizu, T. Kumasaka and S. Baba at the Spring-8 beamline BL41XU for on-site assistance, and N. Yan for discussion and comments on the manuscript. This work was supported by funds from the Ministry of Science and Technology (grant 2009CB918801), National Natural Science Foundation, and Beijing Municipal Commissions of Education and Science and Technology.

Author Contributions D.M., P.L. and Y.S. designed all experiments. D.M., P.L., C.Y., C.F. and P.Y. performed the experiments. D.M., P.L., C.Y., C.F., P.Y., J.W. and Y.S. analysed the data. D.M., P.L., C.Y., J.W. and Y.S. contributed to manuscript preparation. Y.S. wrote the manuscript.

Author Information The atomic coordinates and the structure factor file have been deposited in the Protein Data Bank under accession numbers 4DJK and 4DJL. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.S. (shi-lab@tsinghua.edu.cn).

METHODS

Protein preparation. The cDNA of full-length GadC was subcloned into pET15b (Novagen). Overexpression of GadC was induced by 0.2 mM isopropyl β -D-thiogalactoside (IPTG) when the cell density reached a D_{600} of 1.2. After growth at 37 °C for 4 h, the cells were collected, homogenized in buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl. After further disruption with a French press, cell debris was removed by low-speed centrifugation for 10 min. The supernatant was collected and ultracentrifuged for 1 h at 150,000g. The membrane fraction was collected and homogenized with buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl. N-Octyl- β -D-glucopyranoside (β -OG; Anatrace) was added to the membrane suspension to a final concentration of 2% (w/v) and then incubated for 2 h at 4 °C. After another ultracentrifugation step at 150,000g for 30 min, the supernatant was collected and loaded on Ni^{2+} -nitrilotriacetate affinity resin (Ni-NTA; Qiagen), followed by a wash with 25 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 0.4% n-nonyl- β -D-maltopyranoside (NM; Anatrace). After proteolytic removal of the hexahistidine (His₆) tag on the column, GadC was eluted with buffer containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.4% NM. After concentration to 10–15 mg ml⁻¹, GadC was further purified by gel filtration (Superdex-200 10/30; GE Healthcare). The buffer for gel filtration contained 25 mM Tris-HCl pH 8.0, 150 mM NaCl and various detergents. The peak fractions were collected. The GadC mutants were generated with a standard PCR-based strategy and were subcloned, over-expressed and purified in the same way as the WT protein.

Crystallization. The hanging-drop vapour-diffusion method was performed at 18 °C during crystallization. Crystals belonging to the space group $P2_12_12_1$ were obtained with protein purified in the presence of 0.2% n-nonyl- β -D-glucopyranoside (β -NG; Anatrace) and 0.023% n-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO; Anatrace). The crystallization buffer was 21% PEG400, 100 mM Tris-HCl pH 8.0, 100 mM NaCl, and 325 mM sodium acetate. Rod-shaped crystals appeared overnight and typically grew to full size in about 1 week. Crystals were dehydrated by exposing the drops to air for 5 min. The best diffraction reached 2.95 Å at SSRF beamline BL17U. Platinum derivatives were obtained by soaking the crystals for 48 h in mother liquor containing 10 mg ml⁻¹ $K_2Pt(\text{NO}_2)_4$. Seleno-L-methionine-incorporated crystals were also obtained and reach similar diffraction with heavy-atom-derived crystals. Diffraction data for heavy-atom and selenomethionine derivatives were collected at SPring-8 beamline BL41XU.

Data collection and structure determination. All anomalous diffraction data, including Pt-SAD and SeMet-SAD data, were collected at SPring-8 beamline BL41XU and processed with the package HKL-2000 (ref. 26) with routine procedures. The diffraction images from the severely anisotropic native crystal were collected at SSRF beamline BL17U and integrated with DENZO²⁶. Before the .x files were inputted into SCALEPACK²⁶ for merging and scaling, the anisotropic ellipsoidal truncations on the .x files were performed with the special version of the ellipsoidal truncation program provided by the University of California, Los Angeles, MBI Diffraction Anisotropy Server³¹. The applied resolution limits along the a^* , b^* and c^* directions are 3.77, 3.31 and 3.10 Å, respectively, on the basis of the criterion of F/σ larger than 3.0. The pruned data were then used for structural determination and refinement. Further processing was carried out with programs from the CCP4 suites³². Data collection statistics are summarized in Supplementary Table 1, and Supplementary Table 2 compares the data completeness before and after the anisotropic truncation.

The platinum positions were determined with the program SHELXD³³. The identified platinum sites were then refined, and initial phases were calculated in the program PHASER³⁴. Cross-crystal averaging with all the three data sets combined with solvent flattening, histogram matching and non-crystallographic symmetry (NCS) averaging in DMMulti³⁵ gave a map of sufficient quality for model building. An initial model was built into the experimental map with COOT²⁸. The sequence docking was aided with the selenium sites in the anomalous difference Fourier map. The structure was refined with PHENIX²⁹.

Preparation of oxidized protein. Oxidation of the WT GadC and the GadC mutant L267C/N364C was performed with *o*-phenanthroline copper complex²¹. The oxidation system comprised 25 mM MES buffer pH 5.5, 150 mM KCl, 2.5 mg ml⁻¹ protein and 0.9 mM *o*-phenanthroline copper complex. The reaction was performed for 2 h on ice.

Preparation of liposomes and proteoliposomes. Liposomes of *E. coli* polar lipid (Avanti) was prepared using a standard protocol as described previously³⁶. For the study of membrane potential on substrate transport by GadC, liposomes were loaded with 5 mM GABA and either 120 mM sodium phosphate pH 5.5 or potassium phosphate pH 5.5. For all other transport assays of GadC, liposomes were loaded with various choices of buffer system (25 mM) depending on the assay purposes, 150 mM KCl and 5 mM GABA (or other amino acid and their derivatives). For AdiC, 5 mM GABA was replaced with 5 mM agmatine or arginine. WT or mutant GadC or WT AdiC were incorporated with liposomes to form proteoliposomes by incubation with pre-extruded liposomes together with 1.25% β -OG (Anatrace) at a concentration of 5 µg mg⁻¹ lipids. β -OG was removed by incubation overnight with 400 mg ml⁻¹ Bio-Beads SM2 (Bio-Rad). The proteoliposomes were harvested by ultracentrifugation for 1 h at 150,000g and rinsed twice with resuspension buffer (various choices of buffer system (25 mM), 150 mM KCl). The proteoliposomes were resuspended with the resuspension buffer to a final lipid concentration of 100 mg ml⁻¹.

In vitro transport assay. All transport assays were performed at 25 °C. For the study of membrane potential on substrate transport by GadC, the reaction was initiated by adding proteoliposomes (2 µl) to 100 µl of external buffer containing 120 mM sodium phosphate pH 5.5 or potassium phosphate pH 5.5, 50 µM unlabelled L-glutamic acid and 1 µCi of L-[³H]glutamic acid (specific radioactivity 51.1 Ci mmol⁻¹; PerkinElmer Life Sciences), with or without 1 µg ml⁻¹ valinomycin. For all other L-glutamic acid uptake assays of GadC, the reaction was initiated by adding proteoliposomes (2 µl) to 100 µl of external buffer containing 25 mM pH buffer, 150 mM KCl, 50 µM unlabelled L-glutamic acid and 1 µCi of L-[³H]glutamic acid. The final concentration of L-[³H]glutamic acid in the external buffer was about 0.19 µM. For AdiC, unlabelled and ³H-labelled L-glutamic acid were replaced by unlabelled and ³H-labelled L-arginine. The uptake of ³H-labelled substrate was stopped at the indicated time points by rapidly filtering the reaction solution through a 0.22-µm GSTF filter (Millipore) and washed with 2 ml of ice-cold wash buffer (25 mM glycine pH 9.5, 150 mM KCl). The filter was then taken for liquid scintillation counting. All experiments were repeated at least three times. The reactions lasted for various lengths of time depending on the different assay purposes.

For determination of V_{max} and K_m , the same substrates were used on both sides of proteoliposomes (Glu/Glu, Gln/Gln, GABA/GABA for GadC; Arg/Arg for AdiC). The chosen time points were within the linear range of substrate accumulation. The preparation of proteoliposomes and the transport assay process were as described above.

- Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**, 8060–8065 (2006).
- Bailey, S. The Ccp4 suite—programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
- Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779 (2002).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Cowtan, K. dm: an automated procedure for phase improvement by density modification. *CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38 (1994).
- Veenhoff, L. M. & Poolman, B. Substrate recognition at the cytoplasmic and extracellular binding site of the lactose transport protein of *Streptococcus thermophilus*. *J. Biol. Chem.* **274**, 33244–33250 (1999).

Choice-specific sequences in parietal cortex during a virtual-navigation decision task

Christopher D. Harvey^{1,3,4}†, Philip Coen^{1,4} & David W. Tank^{1,2,3,4}

The posterior parietal cortex (PPC) has an important role in many cognitive behaviours; however, the neural circuit dynamics underlying PPC function are not well understood. Here we optically imaged the spatial and temporal activity patterns of neuronal populations in mice performing a PPC-dependent task that combined a perceptual decision and memory-guided navigation in a virtual environment. Individual neurons had transient activation staggered relative to one another in time, forming a sequence of neuronal activation spanning the entire length of a task trial. Distinct sequences of neurons were triggered on trials with opposite behavioural choices and defined divergent, choice-specific trajectories through a state space of neuronal population activity. Cells participating in the different sequences and at distinct time points in the task were anatomically intermixed over microcircuit length scales (<100 micrometres). During working memory decision tasks, the PPC may therefore perform computations through sequence-based circuit dynamics, rather than long-lived stable states, implemented using anatomically intermingled microcircuits.

In real-world tasks, decision-making and working memory often occur in the context of other complex behaviours, including spatial navigation. For example, when driving through a city towards a destination, sensory information defining context and place engages memory and decision circuits to plan turns at upcoming intersections. The PPC is a prime candidate for the neuronal circuitry combining the cognitive processing elements necessary for such tasks. In primates, the PPC is important for perceptual decision-making and categorization^{1–3}, movement planning⁴ and spatial attention⁵. Studies in rats suggest that the PPC is also important for encoding route progression during navigation^{6–9}. Using a virtual-reality system for mice¹⁰ and cellular resolution optical imaging methods^{11,12}, we developed a T-maze-based navigation task combining all these cognitive elements and characterized the neuronal circuit dynamics in the PPC, which have not been studied in this combined behavioural context.

Neuronal activity patterns in the PPC have been studied using microelectrode recordings during spatial attention, working memory and perceptual decision tasks. These studies have commonly found cells with sustained firing rate changes spanning entire task periods (cue, delay, response periods)^{2,4,5,13}. For example, cells with persistent activity throughout the delay period of memory-guided saccade/reach tasks have frequently been recorded^{14,15}. Also, studies have identified neurons with ramps of increasing firing rate spanning the accumulation of evidence period in a motion perception task^{1,2}. Neurons with sustained activity during the same task period often have similar activity time courses, suggesting the presence of classes of cells (for example, delay cells) and implying that the task-dependent neuronal dynamics are low-dimensional. The low-dimensional dynamics can be reproduced in recurrent attractor network models, in which each cell's activity is typically an amplitude-scaled version of a prototypical time series^{16–21}. In contrast, recent analysis of prefrontal cortex activity has identified heterogeneous neuronal activity time courses, in which a neuron's activity can be thought of as the sum of a few activity modes^{22–24}.

In addition, there is growing evidence for sequences of neuronal activation within local circuits, in which each neuron is active for only a fraction of a task period, including during working memory tasks in the prefrontal cortex^{25–28} and the hippocampus²⁹ and during an object construction task in the PPC³⁰. Sequences suggest dynamics that are high-dimensional, without the presence of classes of cells with relatively homogeneous activity time courses.

We explored whether PPC dynamics are best described in terms of cell classes or high-dimensional dynamics. During a navigation-based decision task, the dynamics were high-dimensional: neurons were active in choice-specific sequences in which information moves from one neuronal population to another across time in the task; although the neuronal activity patterns could be divided into cue, delay and response groups, sequences were present within each group. Furthermore, exploiting the ability of cellular resolution optical measurements to provide the relative anatomical location of the recorded cells^{12,31,32}, we found that neurons active during behaviourally distinct task periods and on trials with different behavioural choices were spatially intermixed over microcircuit length scales.

A PPC-dependent navigation-based decision task

Using a virtual-reality system¹⁰, we trained mice to navigate through a virtual T-maze in a task that incorporated both visual discrimination and a memory-guided response (Methods; Fig. 1a). Visual cues present in the initial section of the maze indicated which direction to turn at the T-intersection to receive a water reward (one set of cues to indicate a right turn and a second set for a left turn; Supplementary Fig. 1c). Between the cue section and the T-intersection, mice ran through a delay maze section that was identical on left and right turn trials; visual information about the reward location was thus present only in the cue section. The task resembled traditional delay tasks^{14,15,33} in its cue-delay-response structure, but differed in that continuing sensorimotor activity was present throughout the task, including the delay period.

¹Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA. ²Bezos Center for Neural Circuit Dynamics, Princeton University, Princeton, New Jersey 08544, USA. ³Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. ⁴Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [†]Present address: Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

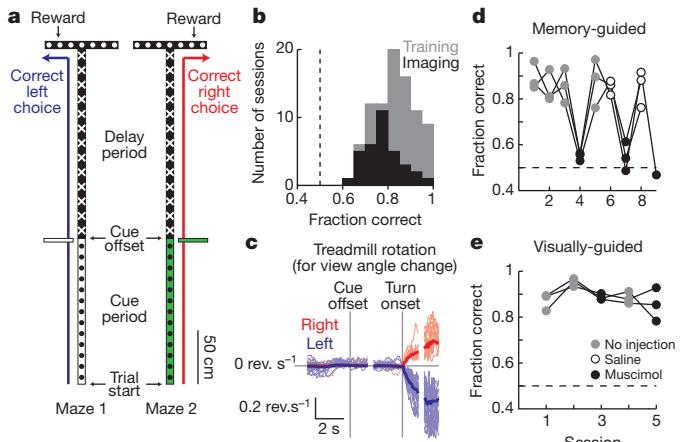


Figure 1 | A PPC-dependent decision task in virtual reality. **a**, Diagram of the two versions of the virtual T-maze that differed only in the cue period and the reward location. Patterns in the diagram reflect the patterns present on the virtual maze walls. **b**, Behavioural performance on individual training (grey) and imaging (black) sessions. **c**, Rotational velocity of the spherical treadmill about the vertical axis for view angle changes on correct right (red) and left (blue) trials, aligned to the cue offset and the turn onset. **d**, Behavioural performance on a memory-guided task from **a** after receiving no injections (grey), saline (open circles) or muscimol (black) bilaterally in the PPC. Connected dots are from individual mice across daily sessions ($n = 3$ mice). **e**, Same as in **d** except for a visually-guided task (Supplementary Fig. 1b; key in **e** applies to **d** also).

Mice performed the task with high levels of accuracy ($83 \pm 9\%$ correct, $P < 0.0001$ versus 50%, *t*-test; Fig. 1b). Within and across trials, individual mice ran at highly consistent speeds (Supplementary Fig. 2); however, running speeds varied between mice, and the maze position at which mice began rotating the spherical treadmill to make a right or left turn differed across trials (turn onset; Fig. 1c, Supplementary Fig. 3; Methods). To compare behaviourally similar epochs of the task across trials, individual trials were aligned to the time points when the cue was no longer visible (cue offset), the turn onset, and the end of the trial; these alignment points formed the boundaries of cue (trial start to cue offset, 5.1 ± 2.6 s), delay (cue offset to turn onset, 4.2 ± 1.2 s) and turn task periods (turn onset to trial end, 3.0 ± 1.8 s).

Because the location of the mouse PPC has not been characterized, we first performed retrograde and anterograde labelling experiments to locate it anatomically (Methods, Supplementary Fig. 4). We identified a region consistent with the rat and primate PPC based on the set of areas from where it received axonal projections, the areas to which it sent projections, and its location relative to other cortical regions^{8,34} (anterior to visual cortex and posterior to somatosensory cortex). We therefore considered this area to be the mouse PPC.

To test if the PPC was required for the behavioural task, we inactivated it using bilateral injections of the GABA_A receptor agonist, muscimol. Muscimol reversibly decreased behavioural performance from high levels of accuracy in control sessions to near chance levels, but did not affect the rate of trials performed (fraction correct; no injections $87 \pm 7\%$, saline $85 \pm 5\%$, muscimol $54 \pm 5\%$ $P < 0.0001$ versus no injections, *t*-test; trials per minute; no injections 3.2 ± 0.5 , muscimol 3.1 ± 0.3 , $P > 0.8$, *t*-test; Fig. 1d). In contrast, PPC inactivation did not significantly affect performance on a visually guided task in which a visual cue was present at the reward site and visible throughout the trial, indicating that the decrease in performance on the memory-guided task was unlikely to be due to a major visual or motor deficit (fraction correct; no injections, $90 \pm 5\%$; muscimol, $87 \pm 5\%$, $P > 0.2$, *t*-test; Fig. 1e, Supplementary Fig. 1b).

Imaging sequences of neuronal activity

We used two-photon microscopy to image layer 2/3 PPC neurons expressing the genetically encoded calcium indicator GCaMP3, which

increases in fluorescence intensity in response to action potential firing³⁵ (Methods). On average, we imaged ~ 65 cells simultaneously within an area $\sim 300 \mu\text{m}$ by $\sim 150 \mu\text{m}$ (range, 37–94 cells). Nearly all imaged cells showed statistically significant Ca^{2+} transients during the behavioural session (96% of cells had >0.2 transients per minute; Methods). Of the cells with high levels of activity ($>2 \text{ Ca}^{2+}$ transients per minute on average; Supplementary Fig. 5), $\sim 73\%$ had significant increases in their mean fluorescence intensity traces ($\Delta F/F$, averaged across trials) during a specific time in the trial or inter-trial interval (task-modulated cells; Fig. 2, Supplementary Figs 6, 7). These task-modulated cells had Ca^{2+} transients for only short time intervals on individual trials ($11 \pm 8\%$ of time points in trials with a transient) such that only a small fraction of neurons was active simultaneously (Fig. 2b–e, Supplementary Figs 8, 9a); cells with prolonged activity patterns covering a large fraction of the trial were not observed. The majority ($\sim 71\%$) of task-modulated cells had significantly different levels of activity on correct right and left choice trials (choice-specific cells). Similar choice-specific, task-modulated activity patterns were observed in extracellular electrophysiological recordings (Supplementary Figs 10, 11). Cells were also active on error trials, such that neurons active during the cue period tended to be correlated with the cue identity, and neurons active during the turn period in general were correlated with the behavioural response (Supplementary Fig. 12). Only a small fraction of cells had obvious reward-related signals ($\sim 2\%$ of active cells with $P < 0.01$, *t*-test, comparing $\Delta F/F$ values within ~ 0.6 s after the reward was given on correct trials or missed on error trials).

When the activity patterns of all the choice-specific, task-modulated cells were ordered according to the time profile of their Ca^{2+} transients, the active periods across cells were staggered relative to one another in time, forming a sequence of neuronal activation covering the entire trial length (Fig. 2c, d, Supplementary Fig. 7c). Different sequences of neurons were activated on left and right trials (Fig. 2c, d and absence of activity in Fig. 2d lower plot, Supplementary Fig. 14e). Although these plots of sequences combined cells from different experiments and averaged across trials, similar properties were observed when considering only the cells imaged in a single mouse and on individual trials (Fig. 2c, Supplementary Fig. 13). Sequences were also apparent in the $\sim 29\%$ of task-modulated cells that did not have choice-specific activity; these cells participated in the sequences for both right and left choice trials (Supplementary Fig. 7b). In total, $\sim 73\%$ of the highly active cells participated in sequences. Sequences similar to those during the task were not observed in shuffled versions of the data set, demonstrating that the sequences were not an artefact created by ordering the data (Supplementary Fig. 14a–d).

Because previous studies of the PPC have categorized cells into classes with cue, delay or response period activity^{14,15}, we examined the activity patterns to see if neurons in the sequence were grouped on the basis of behavioural periods. The distribution of activity times of all cells in the population (calculated for each cell as the centre-of-mass (COM) in time of the mean $\Delta F/F$ during the trial, t_{COM}) had three peaks corresponding to the cue, delay and turn periods, suggesting a possible grouping by behavioural period (Fig. 3a). Consistently, principal component analysis (PCA) of the mean $\Delta F/F$ traces for all cells revealed three intermixed groups, with each group mostly containing cells preferring the same behavioural period (Fig. 3a, b, Supplementary Fig. 15).

Although the population of neurons could be divided into groups, the temporal activity patterns within each individual period were heterogeneous and formed sequences (Fig. 3c). Cells within their preferred period were active for only a fraction of the period ($35 \pm 16\%$ of time points in preferred periods with a Ca^{2+} transient), with different cells active at different times. Although cells with activity covering a large fraction of the period were occasionally observed (for example, Fig. 2c top panel), these cells were rare (4% of cells with activity lasting for $>60\%$ of the period; Supplementary Fig. 9b); the distribution of

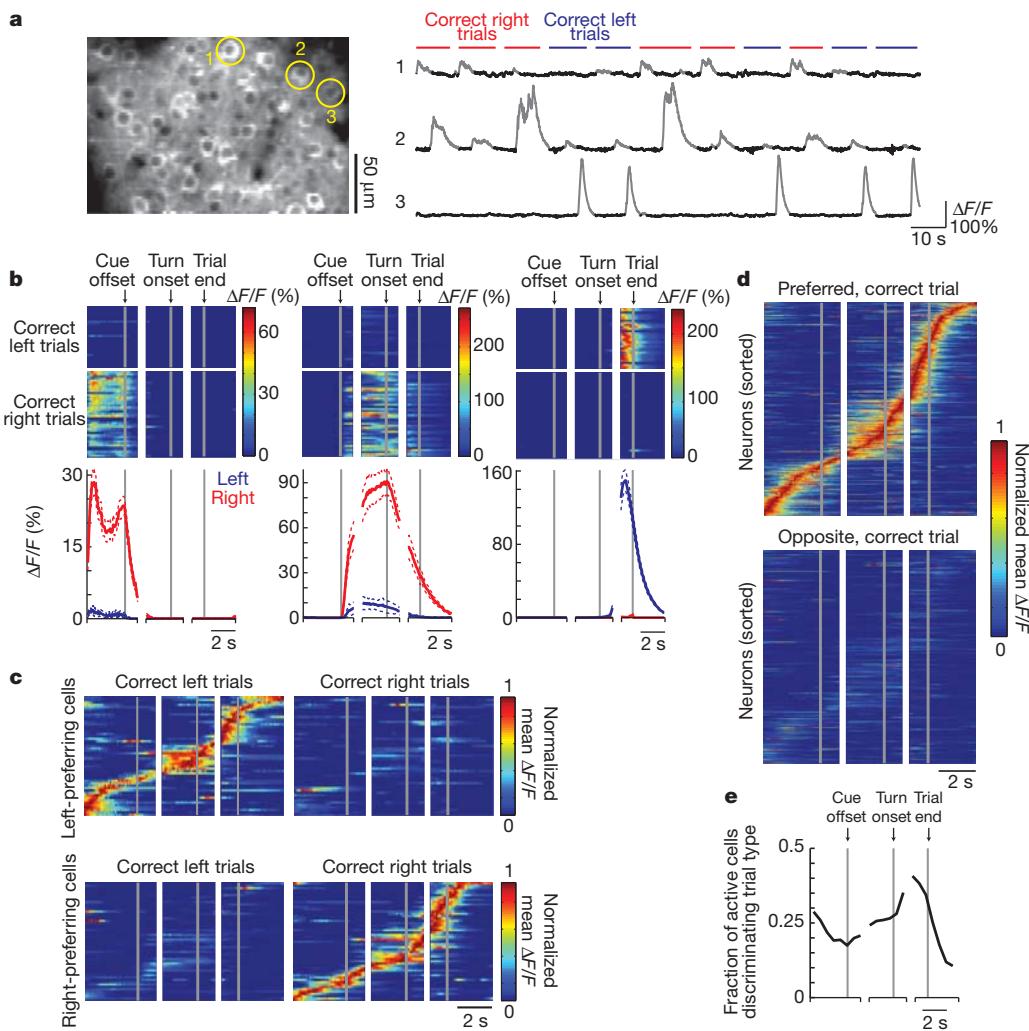


Figure 2 | Imaging PPC neuronal activity during the T-maze task. **a**, Left: example image of GCaMP3-expressing neurons in layer 2/3. Right: example fluorescence intensity traces ($\Delta F/F$; grey portions indicate significant Ca^{2+} transients, Methods) for three example cells from the left panel on correct right (red) and left (blue) trials. **b**, Activity patterns during the task for cells 1–3 from a. Top: colour-coded $\Delta F/F$ traces for individual correct left and right choice trials. Each row is a single trial aligned to the cue offset, turn onset and trial end. Bottom: mean $\Delta F/F$ traces for correct right (red) and left (blue) choice trials. Dashed lines indicate mean \pm s.e.m. **c**, Normalized mean $\Delta F/F$ traces for all the

choice-specific, task-modulated cells (one cell per row) imaged in a single mouse and divided by left-prefering ($n = 51$) and right-prefering ($n = 54$) cells. Traces were normalized to the peak of each cell's mean $\Delta F/F$ trace on preferred trials and sorted by the peak time. Some cells were imaged on different days and in different fields-of-view. **d**, Same as in c, except for all mice ($n = 404$ cells from 6 mice) on preferred and opposite trials. **e**, Fraction of active cells with significantly different activity levels on right and left choice trials as a function of trial time.

epoch coverage by an individual neuron's activity was similar during the delay and cue or turn periods ($P > 0.1$, Kolmogorov-Smirnov test). Furthermore, Pearson's correlations between the non-averaged $\Delta F/F$ traces for cell pairs with the same trial-type and behavioural period preferences varied widely, with a large fraction of pairs having low correlation coefficients (Supplementary Fig. 16a). The low correlation coefficients could be due to activity at different times in the period, as expected for sequences, or activity at the same time in the trial except on different trials. However, the probability that both cells in these pairs had Ca^{2+} transients or that both cells did not have transients on the same trial during their preferred epoch was generally high (Supplementary Fig. 16b), suggesting that the diversity primarily resulted from differences in the activity times of cells within trials. Together these data indicate that classes of cells with homogeneous activity patterns were not present. Rather, choice-specific sequences of neurons were activated in all behavioural periods, with a lower density of cells in the sequence at the borders between periods.

To further examine the sequential neuronal activation on individual trials, we calculated correlations between the activity patterns of cells using the non-averaged $\Delta F/F$ time series. Cells that were active

at similar time points in the trial on average (measured as the difference in t_{COM} values on correct trials, Δt_{COM}) had, on correct trials and in their preferred behavioural periods, highly correlated $\Delta F/F$ traces and peaks in their cross-correlation at a lag approximately equal to Δt_{COM} (Supplementary Fig. 16d–h); these relationships in the non-averaged $\Delta F/F$ time series provide further evidence for sequential activity on individual trials. Cells that were active at similar times on correct trials were also highly correlated on error trials and in non-preferred periods (Supplementary Fig. 16c–f). Additionally, cells that were sequentially activated during their preferred behavioural period on correct trials were sequentially active with similar lags during error trials and other time points in the task (Supplementary Fig. 16g–i).

The choice-specific activity could result if mice experienced different visual stimuli and running patterns on right and left trials and if PPC activity was modulated by those differences. To examine this, we first performed a multiple regression analysis to determine the potential effects of the parameters defining the mouse's running trajectory on the fluorescence changes during the delay period (Supplementary Table 1). These parameters could not explain the choice-specific

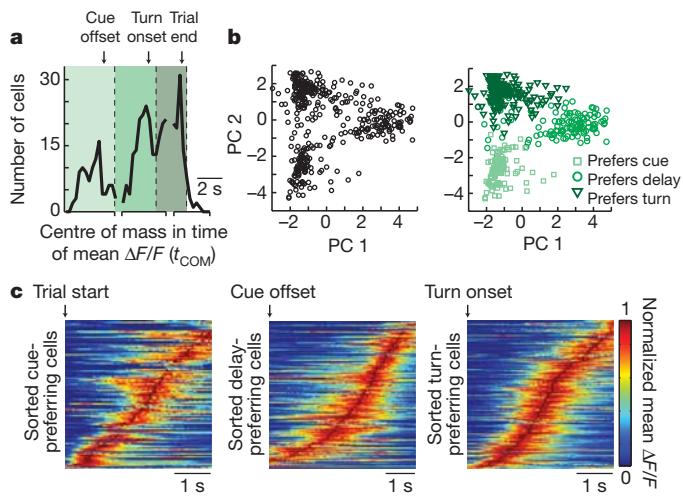


Figure 3 | Neuronal activity in individual behavioural periods. **a**, Histogram of the times of the centre-of-mass of the mean $\Delta F/F$ trace (t_{COM}) for choice-specific, task-modulated cells. Cells were separated into three groups (cue-, delay- and turn-prefering cells; varying shades of green) based on peaks in the distribution. **b**, PCA on the normalized mean $\Delta F/F$ traces for all the choice-specific, task-modulated cells. Left, scores for each cell plotted for the first two principal components (PC 1, 2). Right, cells categorized on the basis of groupings from **a**. $n = 404$ cells. **c**, Sorted normalized mean $\Delta F/F$ traces for cue-prefering ($n = 101$), delay-prefering ($n = 133$) and turn-prefering ($n = 170$) cells, aligned to the trial start, cue offset and turn onset, respectively, on the preferred trial-type.

activity patterns, suggesting that any differences in running trajectories between right and left trials did not trigger the activity we observed. In addition, we performed two sets of experiments to further examine whether the maze visual stimuli alone or the mouse's running patterns triggered PPC activity. In the first experiments, movies of simulated left and right turn runs through the T-maze, which closely approximated real runs, were played to mice that passively viewed the visual scenes (open-loop experiments; Supplementary Fig. 17a, b). In the second experiments, mice were trained on a virtual linear track to perform a simple running back and forth task (run to one end for a reward, turn around, run to the other end for the next reward). The linear track had several visual patterns on the walls, all of which were the same as patterns in portions of the T-maze (Supplementary Fig. 18a, b). The task required similar running and turning behaviours to the T-maze, except that turns were not memory-guided based on visual cues and a delay period, and there was not a two-alternative forced-choice structure. The overall levels of activity during the simulated T-maze runs or during the linear track task were much lower than when mice actively performed the T-maze task (Supplementary Figs 17c, d, 18c, d). Also, only a small fraction of neurons had significant increases in activity at specific locations in the maze, either during the simulated T-maze runs or in the linear track, suggesting that cells were not activated robustly by location-specific visual scenes or running patterns (simulated runs, 1.3% of neurons; linear track, 5.8%; T-maze, 32.3%; Supplementary Figs 17e–k, 18e–k). Together these results suggest that PPC neurons in the T-maze were not activated only by the visual information or by the running patterns of the mouse.

Choice-specific neuronal circuit trajectories

The heterogeneous and sequential neuronal activity patterns during the T-maze task indicated that we should consider the dynamics of the population rather than classes of cells. We therefore analysed the dynamics as a trajectory through a state space of neuronal population activity (neuronal circuit trajectory)^{36–38}. At each time point, the activity state of the circuit containing n simultaneously imaged neurons was defined as a point in an n -dimensional space, with each dimension representing the activity ($\Delta F/F$ values) of a single neuron. Different

trajectories (visualized using factor analysis for dimensionality reduction) were traversed for trials with different behavioural choices (Fig. 4a, b, Supplementary Fig. 19a). The trajectories for correct right and left choice trials began the trial at similar positions, gradually diverged to reach a peak separation near the time of the behavioural choice, and converged to the starting point in the inter-trial interval (Fig. 4a–d). To quantify the trajectory divergence, we used a classifier based on the distance from an individual trial trajectory to the mean right and left choice trajectories at single time points. It was possible, from the activity of a small population of neurons located in close anatomical proximity (~ 65 neurons separated by $< 250 \mu\text{m}$), to predict the mouse's choice on single correct trials at better than chance levels during the cue, delay and turn periods (Methods; Fig. 4e, Supplementary Fig. 19b, c). The activity in the PPC can therefore be considered as divergent, choice-specific trajectories through a state space of neuronal population activity.

Trajectories were highly variable on error trials. Some trajectories began close to the correct choice trajectory during the cue period and

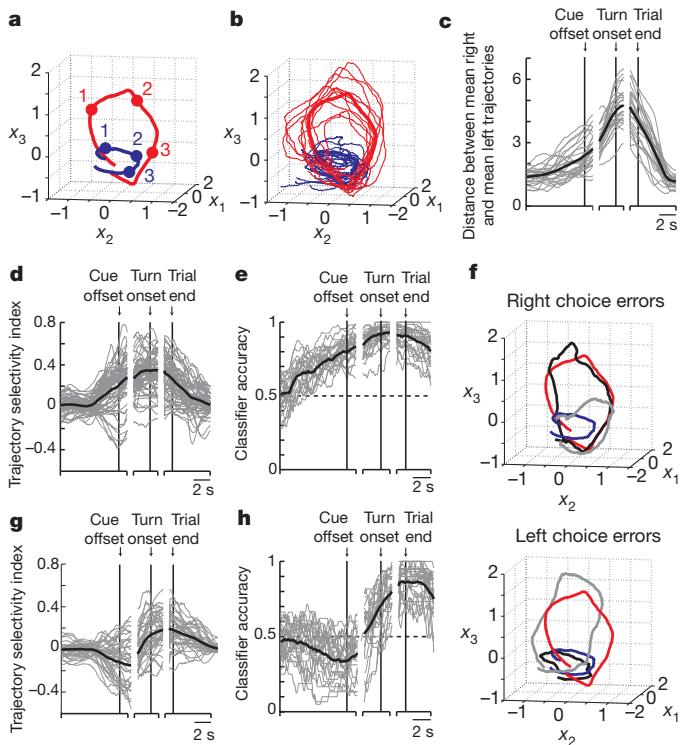


Figure 4 | Neuronal circuit trajectories on correct and error trials. **a**, Time course of mean choice-specific trajectories on correct right (red) and left (blue) choice trials from one session (plotted for the first three common factors). Points labelled 1, 2 and 3 correspond to the mean times of the cue offset, turn onset and trial end, respectively. **b**, Example individual (thin lines) and mean (thick lines) trajectories for correct trials from the session in **a**. **c**, Euclidean distance between the mean trajectories on correct right and left choice trials ($n = 29$ individual sessions, grey). The black line indicates the mean. **d**, Trajectory selectivity index for individual correct trials from a single session, defined on the basis of distances as $(d_{\text{to mean traj, opposite choice}} - d_{\text{to mean traj, same choice}}) / (d_{\text{to mean traj, same choice}} + d_{\text{to mean traj, opposite choice}})$. Values close to 1 and -1 indicate that the individual trial trajectory was near the mean trajectory of the same and the opposite behavioural choice, respectively. Mean trajectories were defined using correct trials only. Grey and black lines indicate individual trials and the mean, respectively. **e**, Classification accuracy for determining the behavioural choice of the mouse at different time points in the task during individual sessions ($n = 29$; black, mean). The classifier was based on a distance-dependent classification scheme (see Methods) using correct trials. **f**, Example individual trial trajectories (grey and black) on right choice and left choice errors trials, plotted with the mean trajectories for correct right (red) and left (blue) choice trials. **g**, Same as for **d**, except on error trials. **h**, Same as for **e**, except on error trials.

transitioned towards the error choice trajectory later in the trial; such transitions occurred at a wide range of points in the trial (grey traces in Fig. 4f, Supplementary Fig. 19f). Other trajectories were similar to the error choice trajectory throughout the trial (Fig. 4f, black traces). On average, the error trial trajectories were more similar to the correct choice trajectory during the cue period and closer to the error choice trajectory during the turn period (cue offset; trajectory selectivity index = -0.13 ± 0.22 , $P < 0.001$ versus 0, t -test: trial end; trajectory selectivity index = 0.18 ± 0.15 , $P < 0.001$ versus 0; Fig. 4g; consistent results based on classification, Fig. 4h). Therefore, individual trajectories transitioned between the mean correct right and left choice trajectories at many time points during trials, but most frequently switched during the delay period.

Anatomical micro-organization of PPC dynamics

To determine how the cells implementing the activity dynamics were anatomically organized, we first examined the neuronal activity patterns to see if cells with different response preferences, such as activity peaks at different times in the trial or different preferences for right and left choice trials, were present in the same regions of the PPC or separated into different areas. Each $\sim 250 \mu\text{m}$ by $\sim 125 \mu\text{m}$ area (that is, field-of-view with simultaneously imaged cells) contained both right and left choice-preferring cells of approximately equal numbers and cells with activity peaks at a wide range of times in the trial (Fig. 5a–c, Supplementary Fig. 20). Next, within each imaged area we compared the activities of pairs of neurons as a function of the distance between the neurons' cell bodies. The difference in the trial-type selectivity for cells in a pair did not depend significantly on the distance between cells, indicating that left and right choice-preferring cells were intermixed ($\rho = 0.04$, Spearman's correlation, $P > 0.05$; Fig. 5e). Similarly, Δt_{COM}

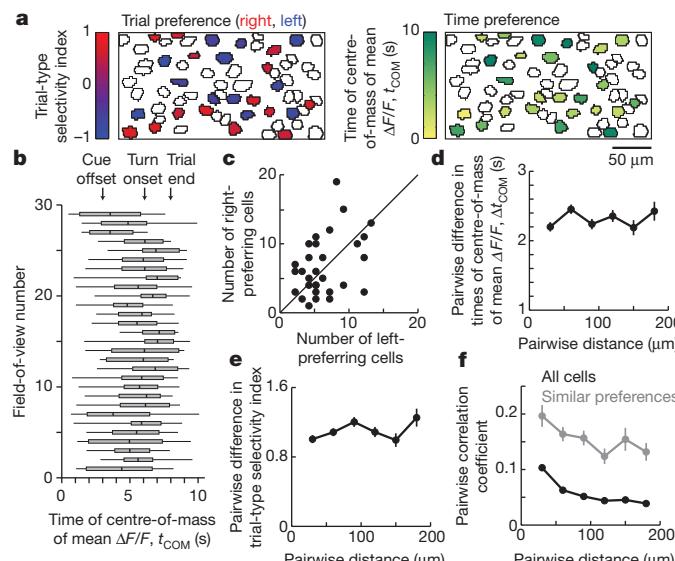


Figure 5 | Anatomical micro-organization in the PPC. **a**, Example field-of-view with cells outlined and choice-specific, task-modulated cells coloured. Left, cells' trial-type selectivity, defined as $(\Delta F/F_{\text{right trials}} - \Delta F/F_{\text{left trials}})/(\Delta F/F_{\text{right trials}} + \Delta F/F_{\text{left trials}})$. Values close to 1 (red) and -1 (blue) indicate right and left choice preferences, respectively. Right, cells' time of the centre-of-mass of the mean $\Delta F/F$ trace (t_{COM}). $t_{\text{COM}} = 0$ corresponds to ~ 3 s before the cue offset. **b**, Box plots of t_{COM} values for task-modulated cells in individual fields-of-view (box edges, first and third quartiles; vertical line in the box, median; whiskers, range). **c**, Number of right and left choice-preferring cells in each field-of-view ($n = 29$; unity line is shown). **d**, Difference in t_{COM} as a function of the distance between cells. **e**, Difference in the trial-type selectivity index as a function of the distance between cells. **f**, Pearson's correlation between non-averaged $\Delta F/F$ traces (all time points) for all pairs of active cells (black) and cell pairs with the same trial-type and behavioural period preference (grey) as a function of the cell–cell distance.

for a cell pair did not differ with the distance between the cells, indicating that cells active during different periods of the task were intermingled ($\rho = -0.01$, $P > 0.6$; Fig. 5d). However, cells that were separated by less anatomical distance had $\Delta F/F$ traces that were significantly more correlated than cell pairs further apart, but the relationship between the correlation coefficient and distance was weak (all pairs of active cells; $\rho = -0.16$, $P < 0.001$; pairs with the same trial-type and behavioural period preference; $\rho = -0.22$, $P < 0.001$; Fig. 5f). This weak relationship could be due to overlapping fluorescence changes, such as from dendritic signals, or could reflect an actual, weak spatial organization. Regardless, cells with highly different activity patterns were intermixed over short length scales, and an anatomical separation of the response properties we measured was not present.

Discussion

The choice-specific sequences of PPC neuronal activation we report here add to the growing list of studies that have identified cortical sequences of activity states during working memory tasks^{27–30}. Furthermore, because we demonstrated that PPC activity was necessary only for the memory-guided task, the sequences of activation were probably important at least for the memory aspect of the task. Sequence-based dynamics may therefore be a common framework for circuit function during memory and decision tasks, including during navigation behaviours. Such dynamics could potentially be implemented using feedforward architectures^{39,40} or liquid state machines^{41,42} related to those that have been proposed for working memory.

Our results also offer a way to unite previous work on neural coding in the PPC. Navigation, memory and choice information may be combined in the sequences such that the identity of the active sequence reflects choice-related information for working memory and movement planning^{4,5}, and that the currently active cell within the sequence reflects spatial or temporal progression through the task^{7,43}, which were highly correlated in our task (Supplementary Fig. 2). It seems unlikely that PPC neurons only provided location information in a context-dependent manner, like hippocampal place cells^{29,44}, because unlike place cells PPC neurons did not encode spatial location during a linear track task (Supplementary Fig. 18) or other tasks^{6,7}.

A possible explanation for heterogeneity and sequences in our experiments versus stereotypy and low-dimensional dynamics, which have been emphasized in previous recordings and models of PPC activity^{1,13–18,20,21}, is that the PPC adopts different dynamics depending on the demands of the behavioural task. Because traditional delayed saccade tasks, for example, have one spatiotemporal component during the delay period (fixation before making a response), the PPC may adopt sustained activity patterns. In contrast, during tasks that involve many spatial and temporal components, as are common in natural behaviours and during navigation, the PPC may utilize sequences of activation. Alternatively, sequences of activity may be present in the primate PPC during traditional tasks but have yet to be identified, consistent with emerging evidence for heterogeneous temporal response properties^{22–24,30}. In addition, different regions or layers of the PPC may have differing activity dynamics⁴⁵, or the dynamics of rodent and primate PPC circuits may differ.

Because cells that were active at distinct time points in the task and that participated in different choice-specific sequences were spatially intermingled, our results indicate that functionally distinct subnetworks are anatomically interlaced in the PPC. This extends previous work in sensory cortex, motor cortex and the hippocampus showing spatial intermixing of heterogeneous response properties in cells encoding qualitatively similar types of information (for example, orientation selectivity in visual cortex) or in cells with activity during similar task epochs^{12,31,32}. Our findings differ from the predictions of models that have emphasized the spatial clustering of similar response patterns, as in functional columns, and that propose connectivity defined by axonal–dendritic overlap without fine-scale specificity^{46,47}. Rather, our results support a model in which microcircuits are formed

by highly specific synaptic connectivity and are composed of neuronal motifs, such as those identified in brain slice recordings in sensory cortices^{48,49} and amongst visual cortical neurons with the same orientation selectivity⁵⁰. Our results showing that cell pairs that were sequentially active during their preferred periods on correct trials had similar activity relationships even during error trials and the inter-trial interval suggest the presence of such motifs (Supplementary Fig. 16g-i).

The behavioural task used here did not isolate the decision-making process. We note however that activity trajectories occasionally switched during a trial between the prototypical correct left and right choice trajectories, including frequently on error trials, suggesting that the mouse's decision was not necessarily irreversibly reached immediately after a trial's start (trajectory selectivity switch during the delay period or last 1 s of the cue period; 63% of error trials, 20% of correct trials; Fig. 4d, g, Supplementary Fig. 19e, f). Sequences may therefore play a role in some aspect of decision-related processes, but further experiments will be necessary to assess this possibility.

Our results motivate consideration of a conceptual framework for decision-making and working memory in which sensory information used for the decision activates a neuronal sequence of activity. The sequence begins in a choice-independent state, which could be mediated by neurons that are not choice-specific (Supplementary Fig. 7b), and then moves towards a choice-specific trajectory and away from other trajectories in a manner dependent on the incoming information. A decision is proposed to be reached when the sequence of activity intersects a choice-specific trajectory; different decisions involve intersections with different trajectories. Upon reaching a decision, a working memory can be maintained by continuing along that choice-specific trajectory. Changing decisions would occur through transitions between trajectories, but as time progresses in the task, the state space distance between trajectories increases, in effect creating a larger barrier to change. In this view, decision-making and working memory utilize an ordered progression through a sequence in which information moves from one population of neurons to another over time. The framework we propose is an extension of a point-of-view first considered in describing the dynamics underlying behavioural choices in the leech nervous system³⁷. It has some similarities with (and some differences from) drift/diffusion-to-bound, race, and recurrent network models of decision-making implemented as neuronal integrator winner-take-all circuits^{2,16}. These models are similar to the trajectory-based view in that different decisions correspond to a divergence in state space surrounding a separatrix. However, these models differ from the sequence framework in that decisions are reached when the activity approaches a choice-specific fixed point with working memory maintained as stable activity at that point. Thus, although these circuits can demonstrate divergent trajectories to reach the fixed points associated with different choices, the trajectories are defined by relatively homogeneous changes in the activity of the population, and the same set of neurons participate in all stages of the decision-making and working memory process.

METHODS SUMMARY

Using a previously described virtual-reality system¹⁰, male C57/BL6 mice were trained using operant conditioning to navigate through a virtual T-maze to receive water rewards. Translation and rotation in the virtual environment were controlled by the mouse's running on a spherical treadmill. Training was performed using shaping implemented as a set of six mazes of increasing task difficulty. Retrograde tracing was performed using fluorescent beads, and anterograde tracing was performed following injections of adeno-associated virus (AAV) containing GFP or GCaMP3. Muscimol injections for PPC inactivation were made bilaterally ~ 350 μ m beneath the dura (50 nl, 1 ng nl^{-1}). Imaging was performed using a custom two-photon microscope incorporated with the virtual-reality system, as described previously¹². Imaging occurred at 2–6 weeks after injection of AAV2/1-synapsin-1-GCaMP3 virus³⁵. A complete description of the experimental methods and data analysis is available in the Supplementary Information.

Received 5 December 2011; accepted 2 February 2012.

Published online 14 March 2012.

- Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
- Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
- Freedman, D. J. & Assad, J. A. A proposed common neural mechanism for categorization and perceptual decisions. *Nature Neurosci.* **14**, 143–146 (2011).
- Andersen, R. A. & Cui, H. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63**, 568–583 (2009).
- Bisley, J. W. & Goldberg, M. E. Attention, intention, and priority in the parietal lobe. *Annu. Rev. Neurosci.* **33**, 1–21 (2010).
- McNaughton, B. L. et al. Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4**, 27–39 (1994).
- Nitz, D. A. Tracking route progression in the posterior parietal cortex. *Neuron* **49**, 747–756 (2006).
- Whitlock, J. R., Sutherland, R. J., Witter, M. P., Moser, M. B. & Moser, E. I. Navigating from hippocampus to parietal cortex. *Proc. Natl. Acad. Sci. USA* **105**, 14755–14762 (2008).
- Calton, J. L. & Taube, J. S. Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. *Neurobiol. Learn. Mem.* **91**, 186–196 (2009).
- Harvey, C. D., Collman, F., Dombek, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
- Dombek, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
- Dombek, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13**, 1433–1440 (2010).
- Curtis, C. E. & Lee, D. Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* **14**, 216–222 (2010).
- Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *J. Neurophysiol.* **66**, 1095–1108 (1991).
- Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* **79**, 2919–2940 (1998).
- Wang, X. J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
- Wong, K. F. & Wang, X. J. A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
- Mazurek, M. E., Roitman, J. D., Ditterich, J. & Shadlen, M. N. A role for neural integrators in perceptual decision making. *Cereb. Cortex* **13**, 1257–1269 (2003).
- Ganguli, S. et al. One-dimensional dynamics of attention and decision making in LIP. *Neuron* **58**, 15–25 (2008).
- Miller, P., Brody, C. D., Romo, R. & Wang, X. J. A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb. Cortex* **13**, 1208–1218 (2003).
- Machens, C. K., Romo, R. & Brody, C. D. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121–1124 (2005).
- Machens, C. K., Romo, R. & Brody, C. D. Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *J. Neurosci.* **30**, 350–360 (2010).
- Jun, J. K. et al. Heterogeneous population coding of a short-term memory and decision task. *J. Neurosci.* **30**, 916–929 (2010).
- Singh, R. & Eliasmith, C. Higher-dimensional neurons explain the tuning and dynamics of working memory cells. *J. Neurosci.* **26**, 3667–3678 (2006).
- Batuev, A. S. Two neuronal systems involved in short-term spatial memory in monkeys. *Acta Neurobiol. Exp. (Warsz.)* **54**, 335–344 (1994).
- Seidemann, E., Meilijison, I., Abeles, M., Bergman, H. & Vaadia, E. Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *J. Neurosci.* **16**, 752–768 (1996).
- Baeg, E. H. et al. Dynamics of population code for working memory in the prefrontal cortex. *Neuron* **40**, 177–188 (2003).
- Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsaki, G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature Neurosci.* **11**, 823–833 (2008).
- Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008).
- Crowe, D. A., Averbeck, B. B. & Chafee, M. V. Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *J. Neurosci.* **30**, 11640–11653 (2010).
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
- Komiyama, T. et al. Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464**, 1182–1186 (2010).
- Erlich, J. C., Bialek, M. & Brody, C. D. A cortical substrate for memory-guided orienting in the rat. *Neuron* **72**, 330–343 (2011).

34. Corwin, J. V. & Reep, R. L. Rodent posterior parietal cortex as a component of a cortical network mediating directed spatial attention. *Psychobiology* **26**, 87–102 (1998).
35. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
36. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).
37. Briggman, K. L., Abarbanel, H. D. & Kristan, W. B. Jr. Optical imaging of neuronal populations during decision-making. *Science* **307**, 896–901 (2005).
38. Churchland, M. M., Yu, B. M., Sahani, M. & Shenoy, K. V. Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Curr. Opin. Neurobiol.* **17**, 609–618 (2007).
39. Goldman, M. S. Memory without feedback in a neural network. *Neuron* **61**, 621–634 (2009).
40. Ganguli, S., Huh, D. & Sompolinsky, H. Memory traces in dynamical systems. *Proc. Natl. Acad. Sci. USA* **105**, 18970–18975 (2008).
41. Maass, W., Joshi, P. & Sontag, E. D. Computational aspects of feedback in neural circuits. *PLOS Comput. Biol.* **3**, e165 (2007).
42. Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**, 544–557 (2009).
43. Leon, M. I. & Shadlen, M. N. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* **38**, 317–327 (2003).
44. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
45. Burke, S. N. *et al.* Differential encoding of behavior and spatial context in deep and superficial layers of the neocortex. *Neuron* **45**, 667–674 (2005).
46. Braintenberg, V. B. & Schuz, A. *Anatomy of the Cortex: Statistics and Geometry* (Springer, 1991).
47. Binzegger, T., Douglas, R. J. & Martin, K. A. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24**, 8441–8453 (2004).
48. Yoshimura, Y., Dantzker, J. L. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).
49. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
50. Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Dombeck for assistance with imaging methods and analysis; C. Domnisoru, M. de Bettencourt, C. Brody and A. Miri for discussions; and M. Goldman, J. Hopfield, D. Aronov, B. Scott and T. Hanks for comments on the manuscript. This work was supported by the NIH (R01-MH083686; R01-NS068148), a fellowship from the Helen Hay Whitney Foundation (C.D.H.), and a Burroughs Wellcome Fund Career Award at the Scientific Interface (C.D.H.).

Author Contributions C.D.H. performed experiments with assistance from P.C. on the retrograde tracing experiments; D.W.T. implemented the imaging instrumentation; C.D.H. analysed the data with strategy and methods contributions from D.W.T.; C.D.H. and D.W.T. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.D.H. (christopher_harvey@hms.harvard.edu) or D.W.T. (dwtank@princeton.edu).

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature10976

Corrigendum: Functional complementation between FADD and RIP1 in embryos and lymphocytes

Haibing Zhang, Xiaohui Zhou, Thomas McQuade, Jinghe Li, Francis Ka-Ming Chan & Jianke Zhang

Nature **471**, 373–376 (2011)

In Fig. 3b of this Letter, the plot for the NP205 epitope response in the *Fadd*^{−/−}*Rip1*^{−/−} double-knockout (DKO) splenocytes was a duplicate of that for the NP38 epitope response. The correct figure panel is shown below. The panel has been corrected in the HTML and PDF versions online.

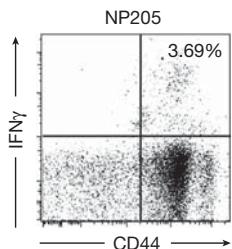


Figure 1 | This is the correct panel from Fig. 3b of the original Letter.

VIROLOGY

Influenza's tale of tails

Epigenetics is a hot new research field, but it seems that the influenza virus already has it figured out. By mimicking epigenetic regulation in human cells, one flu strain suppresses the expression of antiviral genes.

ALEXEI L. KRASNOSELSKY
& MICHAEL G. KATZE

Influenza virus infections are sometimes severe, or even deadly, but most people recover within a few days, helped by their body's immune system. Yet this apparently simple interaction between the virus and its host is deceptive. Influenza viruses are, in fact, masters at circumventing their hosts' defences — they co-opt cellular protein-synthesis pathways to produce viral proteins, for example, and antagonize aspects of the immune response. In a paper published on *Nature*'s website today, Marazzi *et al.*¹ describe another viral evasion tactic. They show that an influenza protein called NS1 mimics a host-cell histone protein that is involved in regulating gene expression. By hijacking this regulatory machinery, the virus inhibits the cell's production of antiviral proteins.

Cells tightly regulate the expression of specific genes according to an organism's developmental stage and physiological state, and in response to environmental stimuli such as infection. Gene expression can be altered in various ways, including by transcription-factor proteins, which bind to DNA and regulate the transcription of certain genes, or through epigenetic modification — heritable chemical changes to DNA and associated proteins that do not affect the DNA sequence but alter the rate of transcription. Epigenetics has rapidly become a much-studied topic since these regulatory mechanisms were first reported around 20 years ago.

A common epigenetic alteration is the addition of methyl or acetyl groups to histone proteins, which are the 'building blocks' of chromatin, the scaffold that packages DNA into chromosomes. The four main histone proteins (H2A, H2B, H3 and H4) consist of a globular domain and an unstructured amino-terminal tail region that is dominated by the positively charged amino acids lysine and arginine. These amino acids can be modified by enzymes that add or remove methyl or acetyl groups². Such modifications alter the histone-DNA interaction, which in turn leads to dynamic changes in chromatin that allow specific areas of the chromatin to unfold, or other proteins to be recruited to the site. Thus,

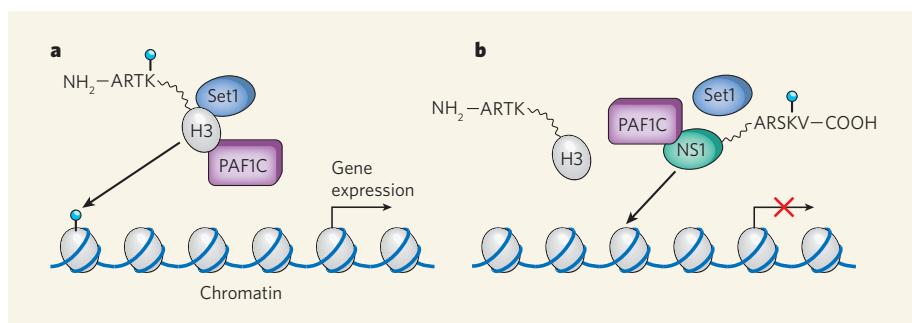


Figure 1 | Interfering influenza. Histones are the building-block proteins around which DNA is wound to form chromatin. Epigenetic changes to histones, such as the addition of methyl groups (blue circles), can affect the rate of gene transcription at specific DNA sequences. **a**, The amino-terminal (NH₂) 'tail' of the human H3 histone protein contains a string of amino acids (ARTK) that can be methylated at the lysine (K) residue by a methyltransferase enzyme, Set1. This enzyme is recruited by a protein complex called PAF1C, which also promotes elongation of RNA molecules during gene transcription. Methylation of the lysine in ARTK of the H3 histone tail is associated with active gene transcription. **b**, Marazzi and colleagues¹ show that the NS1 protein of the H3N2 influenza virus contains an amino-acid sequence (ARSK, close to the protein's carboxy (COOH) terminus) that mimics the H3 ARTK sequence. Using this sequence, viral NS1 can bind directly to PAF1C, which might direct Set1's methylation activity to NS1 rather than to the H3 histone. The authors also show that NS1 is deposited at the promoter regions of the host's DNA to which methylated H3K4 would typically bind, and that this interference suppresses the expression of antiviral genes.

by regulating the accessibility of a particular region of DNA, epigenetic modifications can regulate gene expression.

Some epigenetic changes are generally associated with a site of active gene transcription³. An example of such a change is methylation at the lysine residue (K) of the amino acid sequence designated ARTK in histone H3 (known as the H3K4 modification). Marazzi *et al.*¹ found that the H3N2 influenza virus contains an amino-acid sequence (ARSK) very similar to the histone's ARTK sequence (Fig. 1). This mimic sequence is found in the virus's NS1 protein, which is not essential for viral structure but is known to have other roles in evading the host's immune system⁴. The authors also show that this sequence similarity is functional — the NS1 tail can serve as a substrate for the histone-modifying enzyme Set1, a lysine methyltransferase.

Marazzi and colleagues also demonstrate that H3N2's histone-mimic tail binds directly to the transcription-elongation complex PAF1C (Fig. 1). PAF1C accompanies the transcription enzyme RNA polymerase II during the formation of messenger RNA, and is thought to recruit the Set1 enzyme that leads

to H3K4 methylation⁵. The authors propose that NS1 binding to PAF1C interferes with gene transcription in the host cell. Indeed, the researchers found that when cells from human lung tissue were infected with the H3N2 virus, transcription of rapidly inducible genes was halted. But when the cells were infected with H3N2 viruses that had been mutated so that they could no longer bind PAF1C, transcription of these genes was unaffected. Furthermore, the authors provide compelling evidence that it is specifically the transcription of antiviral genes that is interfered with by the interaction between NS1's histone-like tail and PAF1C. Thus, histone mimicry seems to support viral infection, and may give a selective advantage to viruses that have this ability.

Is this mechanism unique to H3N2 viruses, and could it be a recent evolutionary adaptation? Moreover, does possession of this mechanism have a role in the variation in infectivity and disease severity that is seen among influenza strains? The H3N2 virus is an emerging influenza subtype that has caused only a limited number of human infections during the 2011–12 flu season. However,

they seem to cause illness of around the same severity as that induced by standard seasonal influenza viruses, which do not contain the NS1 histone-mimic tail¹. It will be necessary to assess other emerging influenza strains for histone mimicry to determine whether this mechanism is evolutionarily new, and to study more influenza cases to predict its potential impact on public health. Even if histone mimicry is not widespread and does not dramatically alter disease severity, Marazzi and colleagues' discovery may provide a novel target for antiviral drugs aimed at modulating the host's antiviral response.

The finding raises another intriguing question: what would happen if a person is concurrently infected with H3N2 virus and a more virulent influenza virus, such as the highly pathogenic avian H5N1? Or what

might occur if an H5N1 virus acquires the NS1 sequences required for the protein to function as a histone mimic? The combined effect of a suppressed antiviral response and a highly virulent virus might lead to devastating consequences for the host. Alternatively, if an overly active host response is itself the main culprit in causing severe influenza, as has been suggested⁶, a suppressed antiviral response could actually reduce disease pathology.

Research aimed at addressing these questions may be hampered by an increasing wariness about generating recombinant H5N1 viruses⁷. But ongoing influenza research remains essential. Although epigenetic gene regulation is new to scientists, other mechanisms of gene regulation no doubt remain to be discovered, and we can be certain that viruses have already evolved to interfere with these

regulatory processes, and to use them to their advantage. ■

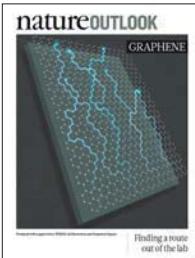
Alexei L. Krasnoselsky and Michael G. Katze
are in the Department of Microbiology and
Washington National Primate Research
Center, University of Washington, Seattle,
Washington 98195-8070, USA.
e-mails: akrasnos@u.washington.edu;
honey@u.washington.edu

1. Marazzi, I. et al. *Nature* <http://dx.doi.org/10.1038/nature10892> (2012).
2. Berger, S. L. *Nature* **447**, 407–412 (2007).
3. Ng, S. S., Yue, W. W., Oppermann, U. & Klose, R. J. *Cell Mol. Life Sci.* **66**, 407–422 (2009).
4. Garcia-Sastre, A. *Virus Res.* **162**, 12–18 (2011).
5. Jaehning, J. A. *Biochim. Biophys. Acta* **1799**, 379–388 (2010).
6. Kash, J. C. et al. *Nature* **443**, 578–581 (2006).
7. Kawaoka, Y. *Nature* **482**, 155 (2012).

nature OUTLOOK

GRAPHENE

15 March 2012 / Vol 483 / Issue No. 7389



COVER ART: NIK SPENCER

Editorial

Herb Brody, Michelle Grayson, Tony Scully, Nick Haines

Art & Design

Wes Fernandes, Alisdair Macdonald, Andrea Duffy

Production

Karl Smart, Emilia Orviss, Leonora Dawson-Bowling

Sponsorship

Patrick Murphy, Yvette Smith, Gerard Preston

Marketing

Elena Woodstock, Hannah Phipps

Project Manager

Christian Manco

Art Director

Kelly Buckheit Krause

Magazine Editor

Tim Appenzeller

Editor-in-Chief

Philip Campbell

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol XXX, No. XXXX Suppl, Sxx–Sxx (2011). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Graphene* supplement can be found at <http://www.nature.com/nature/outlook/graphene/>

Historical epochs can be defined by materials: from the Stone Age, Bronze Age, and Iron Age to the modern eras of steel and, for the last 60 years, silicon. These are the stuff that defined whole civilizations. And we may be on the threshold of a time dominated by one of the most common of all elements: carbon.

The 2010 Nobel Prize in Physics went to scientists who revealed the remarkable properties of graphene, a single layer of carbon in which the atoms are arranged in hexagons. Graphene is the latest allotrope of carbon to be under the spotlight, following from the fullerenes ('buckyballs') of the 1980s and the nanotubes of the 1990s.

Talk to some of the scientists studying graphene and they can find it hard to contain their excitement — it's as if they've been handed a new building block from which to engineer the world. Electric charge moves through graphene impossibly fast. It's extremely strong, yet flexible. Add to that its impressive optical behaviour — it's equally transparent to ultraviolet, visible and infrared light — and you whet the appetites of those developing diverse technologies, including computer chips, touch screens and solar cells (page S38). On the downside, however, it is difficult to produce graphene in pieces large enough for practical purposes (page S32).

Although graphene cannot directly replace silicon in the digital switches that comprise computers (page S34), comparing it to silicon (page S43) is a useful analysis, and serves as a humbling reminder of how tricky it is to predict what any new material's greatest impact will be.

We acknowledge the financial support of POSCO, LG Electronics and Graphene Square in producing this Outlook. As always, *Nature* takes full responsibility for all editorial content.

Herb Brody

Supplements Editor, Nature Outlook.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas — including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

Feedback@nature.com
Copyright © 2012 Nature Publishing Group

CONTENTS

S30 MATERIALS SCIENCE

Super carbon

What is so special about graphene?

S32 PRODUCTION

Beyond sticky tape

Industry in need of large pristine sheets

S34 ELECTRONICS

Back to analogue

Making the most of its intrinsic qualities

S37 BIOELECTRONICS

The bionic material

At the interface of biological sensing

S38 OPTOELECTRONICS

Come into the light

Its brilliance is becoming transparent

S40 Q&A

Taking charge

Tomás Palacios, head of a new specialist centre at MIT, explains the challenges facing this fledgling technology

S42 PERSPECTIVE

A means to an end

Rodney Ruoff

S43 MATERIAL HISTORY

Learning from silicon

Are there lessons to be learned from silicon?

COLLECTION

S45 Electronic transport in polycrystalline graphene

Y. Oleg V. Yazyev and Steven G. Louie

S49 Graphene photonics and optoelectronics

F. Bonaccorso *et al.*

S61 High-frequency, scaled graphene transistors on diamond-like carbon

Wu *et al.*

S66 Tunable metal-insulator transition in double-layer graphene heterostructures

L.A. Ponomarenko *et al.*

S70 Roll-to-roll production of 30-inch graphene films for transparent electrodes

S. Bae *et al.*

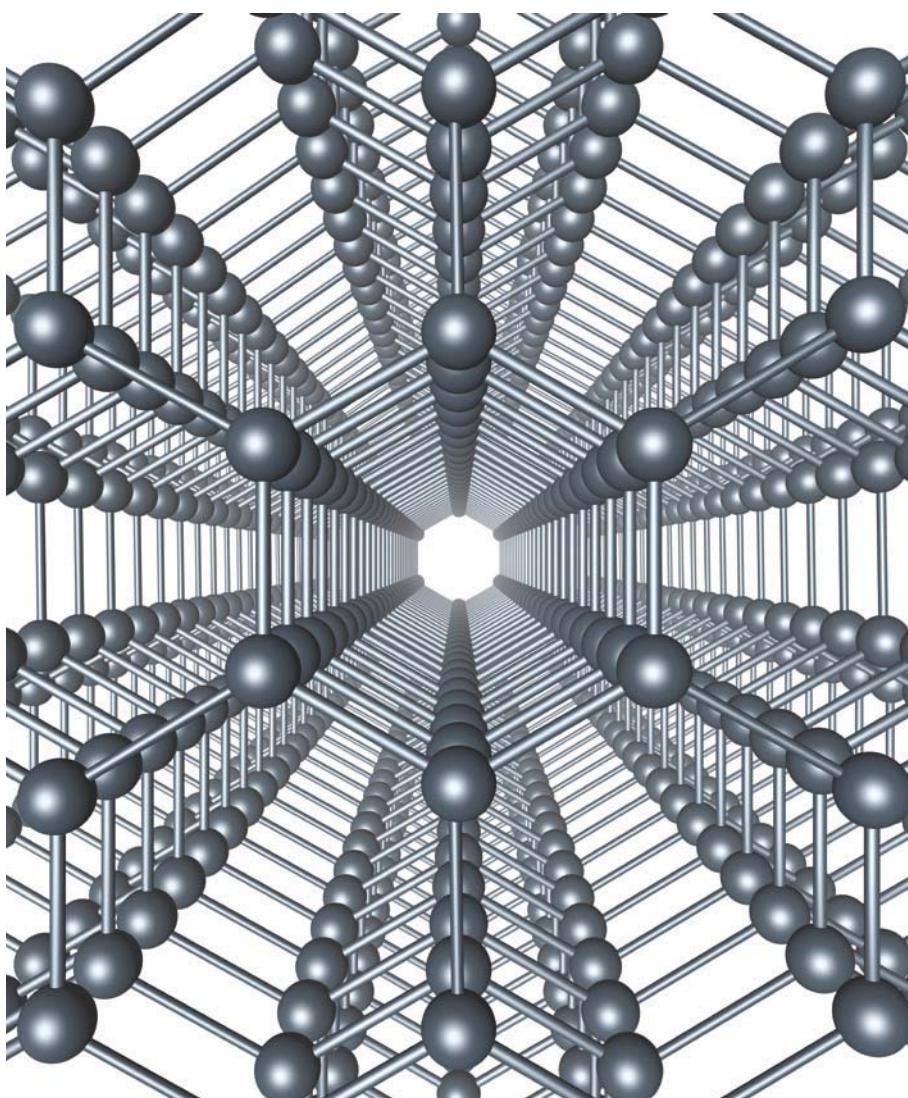
use of sticky tape, was a sheet of carbon a few micrometres across and just a single atom thick. This is because in graphite the atomic bonds between the layers are much weaker than the bonds across its layers.

Graphene's arrangement gives rise to some astonishing properties. Electrons move through it as if they have no mass. The material boasts an electron mobility (a measure of how fast charges travel) 100 times that of silicon. At room temperature, electrons can travel for several micrometres through graphene without scattering — an order of magnitude further than in any other material. And researchers have demonstrated graphene transistors operating at more than twice the speed of the best silicon transistors of similar size. The behaviour of electrons in graphene also provides physicists studying quantum mechanics with a new research tool. Instead of having to use superconductors or fluids cooled to extremely low temperatures, they can observe the seemingly massless activity of electrons on their lab benches.

Graphene stands out in other ways as well. Although it is the thinnest material known to exist, it is also the strongest ever measured: 100 times stronger than steel. It conducts heat better than the previous champion, diamond. The atoms in graphene are packed so tightly that not even the smallest atom — helium — can pass through it. Yet graphene also stretches easily, giving it the ability to bend where other hardy materials would snap.

One property that graphene lacks is a bandgap — a range of energy states where electrons cannot exist. Bandgap is a crucial characteristic that enables semiconductor transistors to switch between on and off states to form the zeroes and ones of digital technology. But what's a problem for electronics is a boon for optics: graphene's lack of a bandgap means the material absorbs light at roughly the same level across the spectrum, from ultraviolet far into infrared. Exploiting graphene's optical and electronic abilities along with its strength and flexibility could lead to foldable plastic smartphones, cheaper solar cells, or sensors that can detect single molecules of gas or identify individual DNA bases.

Despite the intense research interest in graphene, real-world applications are in their infancy. It's not yet clear if it will be able to supplant silicon in transistors any time soon (see 'Back to analogue', page S34). And although carbon is ubiquitous — simply heating any organic material to high temperature in a vacuum produces plenty — engineers are still grappling with ways to make high-quality graphene in sufficient quantities and at a cost to sustain commercial applications (see 'Beyond sticky tape', page S32). Silicon took decades to find its signature role in technology; graphene's journey has only just begun. ■



MATERIALS SCIENCE

Super carbon

Graphene is phenomenally strong, thin, flexible, transparent and conductive — and applications beckon.

BY NEIL SAVAGE

One day in 2004, Andre Geim and Konstantin Novoselov, two chemists at the University of Manchester in the United Kingdom, peeled down graphite to a speck of graphene and ran electrons through it. In doing so they opened up a world of possibilities based on this material's remarkable properties — and won themselves the 2010 Nobel Prize in Physics.

Though only recently identified as a distinct form of carbon, graphene is not exotic. Layers of it stacked together form graphite, the 'lead'

in pencils. Graphene rolled up becomes a carbon nanotube. Wrapped into a sphere, it's a fullerene or, colloquially, a 'buckyball'.

Unlike other carbon forms, graphene is a two-dimensional material, and the first example of such a thing in the real world. Philip Russell Wallace, a theoretical physicist at McGill University in Montreal, Canada, predicted the electronic structure of graphene back in 1947, but many scientists believed 2D crystals were too unstable to exist. What Geim and Novoselov managed to pull off, through their savvy

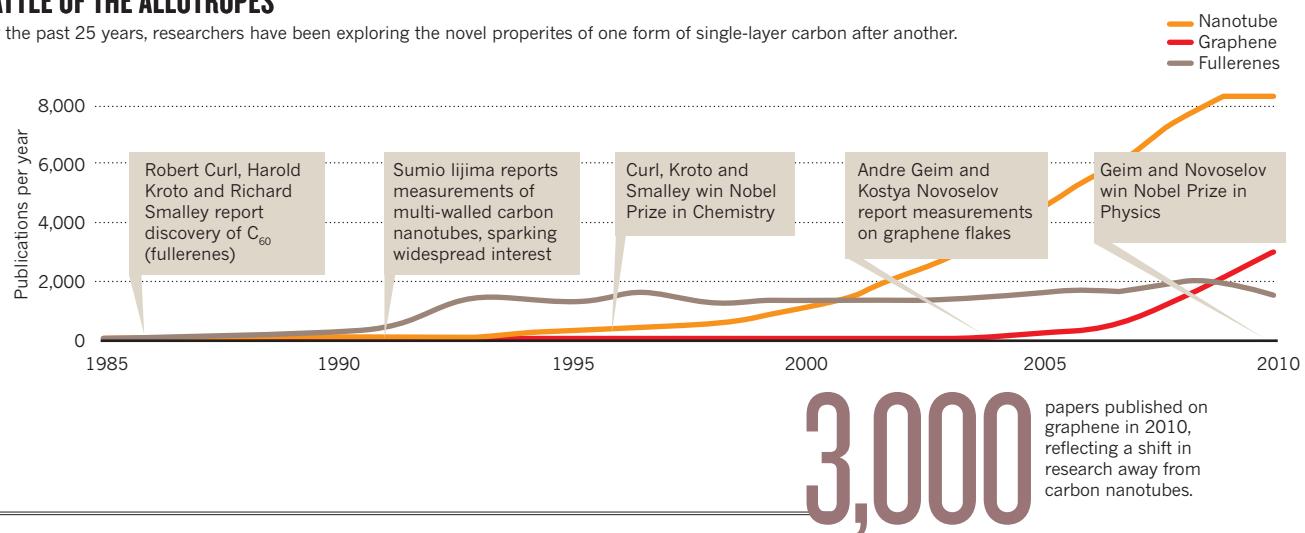
Neil Savage is a freelance writer based in Lowell, Massachusetts.

NEW STAR OF THE MATERIAL WORLD

BATTLE OF THE ALLOTROPIES

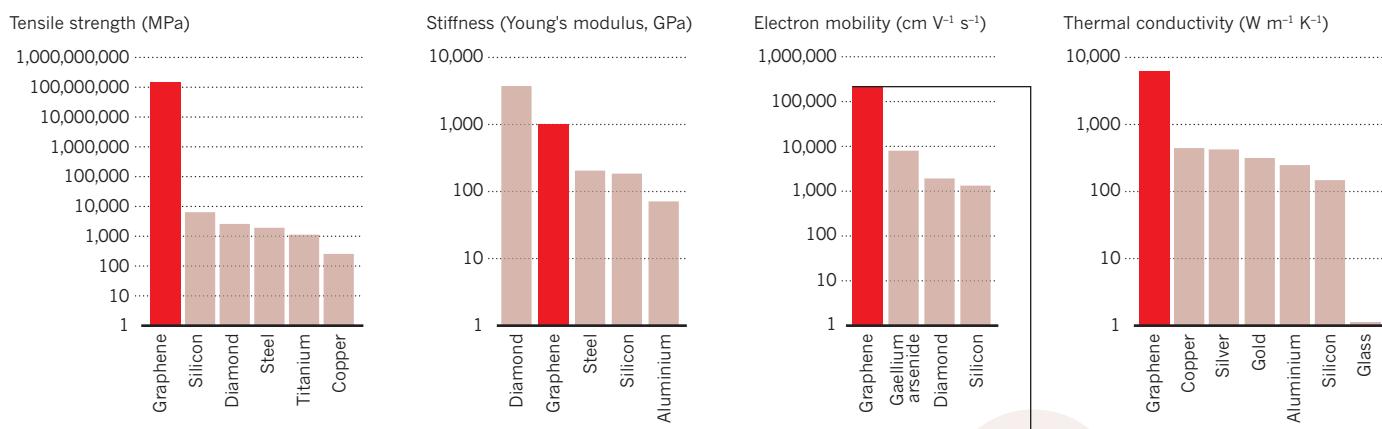
For the past 25 years, researchers have been exploring the novel properties of one form of single-layer carbon after another.

THOMSON REUTERS WEB OF SCIENCE



SUPER-MATERIAL

Graphene stands out for its superlative mechanical, thermal and electronic properties.



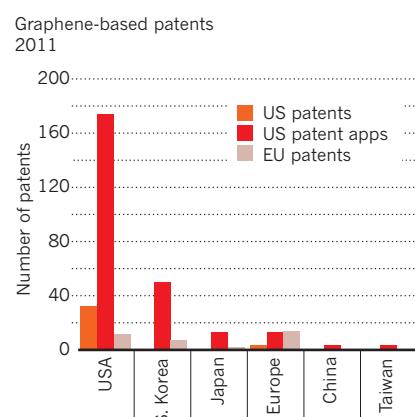
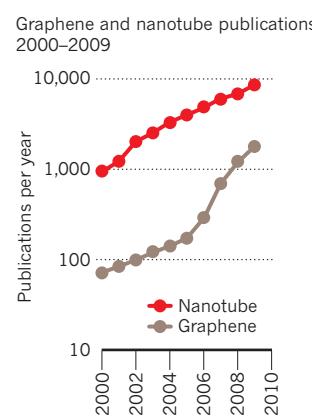
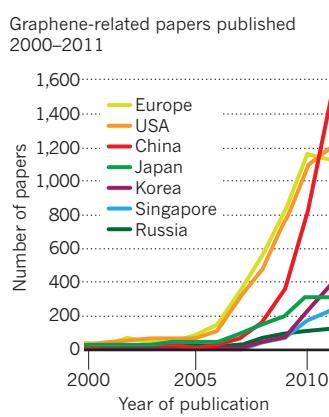
200K

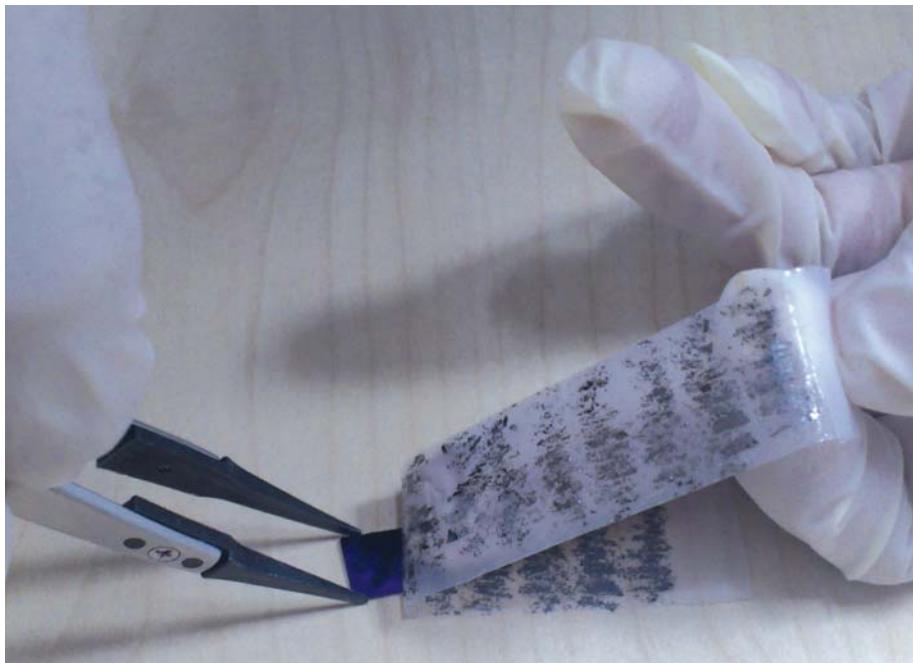
Graphene's extraordinarily high electron mobility could make it suitable for super-fast devices.

EXPLOSIVE GROWTH

As research on graphene takes off, technology development is heavily concentrated in the United States and South Korea.

BYUNG HEE HONG/KOREA GRAPHENE RESEARCH SOCIETY





PRODUCTION

Beyond sticky tape

Flecks of graphene are easy to make. But producing sheets of pristine, electronics-quality material is another matter.

BY RICHARD VAN NOORDEN

Unlike many hi-tech materials, graphene is so simple to make that anyone can produce it at home. First, press adhesive tape onto a chunk of graphite and pull: this peels off a thin flake of grey-black carbon. Then repeatedly stick the carbon-covered tape against itself and peel away: the carbon flake breaks up further into thin, faint fragments, each hundreds of micrometres across. After a few rounds of this, some flakes have been whittled away to single-atom thickness — a fleck of graphene.

All in all, it is a pleasingly artisanal way to make a remarkable material. The graphene is high quality, too: it was on similar products that Manchester University physicists Andre Geim and Kostya Novoselov measured extraordinary electron-transport properties in 2004 — experiments for which they won the 2010 Nobel Prize in Physics. “You can make several square millimetres with a couple of hours’ work,” says Peter Blake, cofounder of Manchester-based Graphene Industries, which sells such flakes to researchers for £500–1000 (US\$780–1550) each.

But one can’t mass produce graphene with sticky tape. To do this, researchers use chemicals to split graphite into graphene platelets by the tonne. Unfortunately, with speed and volume comes variable quality: few of these platelets are

perfect single sheets. The platelets are typically rough stacks of 5 to 10 carbon layers riddled with defects, and begrimed with other chemicals. Yet they still have the attributes that make graphene so appealing: they’re lightweight, strong, have a large surface area and conduct electricity and heat. It’s already possible to buy dozens of different grades of graphene made this way as powders or in solution. Developing applications range from conductive inks and fillers in composites, to battery electrodes and electrochemical sensors.

For more demanding electronics, however, researchers need much purer graphene in large sheets — millimetres or centimetres in diameter. The aim is to grow theoretically perfect graphene: a hexagonal, honeycomb carbon lattice with not a single atom out of place. In practice, a real carbon lattice is strained by defects, polluted by 5- or 7-membered rings, its edge is ragged, and it sits on a substrate that interferes with its properties. These imperfections impair conductivity and other electronic properties. Whether that matters in practice depends on the purpose: transparent conducting films in touch screens, for example, need not be as conductive as those in solar cells.

Large carbon sheets and small imperfect carbon platelets are such contrasting forms of graphene that they could almost be considered different materials. In fact, each version has its merits, depending on cost and application. “We

don’t all buy \$5,000 suits. A lot of the time a \$300 suit is all you need to wear,” says James Tour, an organic chemist who specializes in nanotechnology at Rice University in Houston, Texas. But for materials scientists allured by graphene’s extraordinary potential in electronics and optics it is the \$5,000 suit — or high-quality graphene — that they strive to produce on a large scale.

EXFOLIATING GRAPHITE

For at least a century, researchers have been producing stacks of thin carbon platelets by pulling graphite apart into its constituent layers. (A description of a suspension of small graphite flakes was published in 1907, for example). But these days manufacturers, alert to graphene’s potential, produce consistently thinner platelets. Graphite’s layers are very close-knit, so acids are used to oxidize the material to graphite oxide, layers of which are more easily separated to form graphene oxide. Chemical reduction of graphene oxide leaves graphene platelets crumpled in a solution, or as a powder: a process first reported in 1961 by German chemist Hanns-Peter Boehm, who called his product ‘thinnest carbon films’. It is this method — with proprietary modifications — that companies such as Vorbeck Materials, in Jessup, Maryland; Angstrom Materials in Dayton, Ohio; and XG Sciences in Lansing, Michigan, still use to produce large quantities of graphene. Researchers constantly invent twists in the technique: in 2011, Chao Gao at Zhejiang University in Hangzhou, China, reported spinning a concentrated solution of graphene oxide flakes into fibres several metres in length, which could then be chemically reduced to filaments of pure graphene¹.

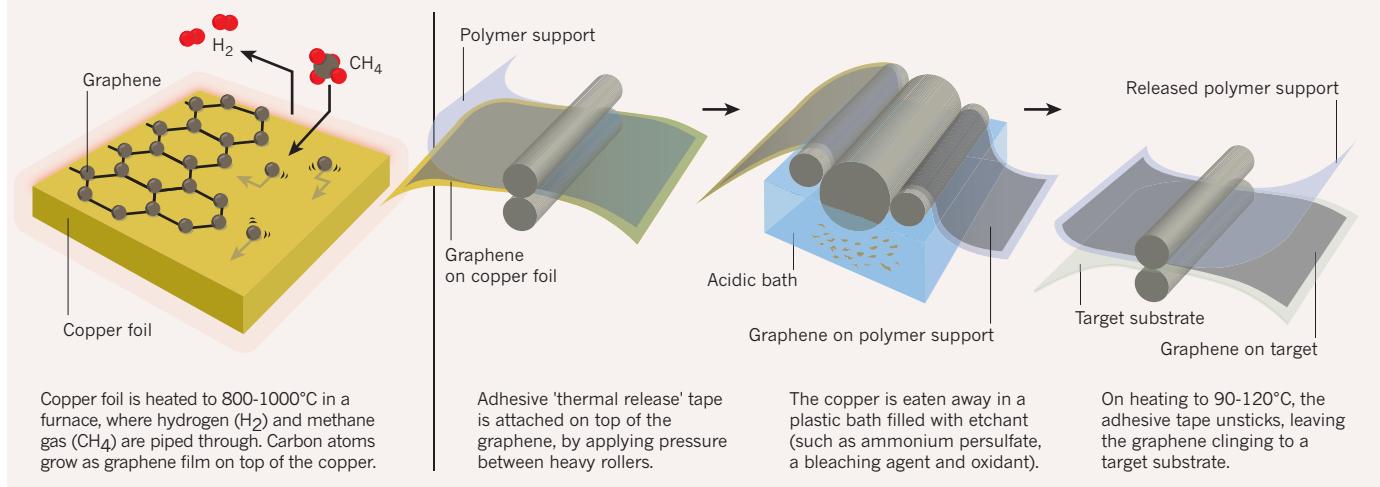
Angstrom scientists have found a way to produce graphene without the harsh acid treatment of the graphite oxide route. They disperse pieces of graphite in water and surfactant, and use ultrasound to peel off layers of carbon: a more pristine graphene as fewer chemicals are involved in its production. Andrea Ferrari, professor of nanotechnology at the University of Cambridge, UK, who has worked on solutions of graphene for ink-jet printing, reckons this method might be useful in electronics that do not need the highest quality graphene, such as touch screens, ‘smart windows’ that can adjust light levels when a voltage is applied, or conductive fabrics that might carry signals from iPods, for example, or heart monitors.

GROWING GRAPHENE

To make larger, higher-quality sheets of graphene, researchers are growing the material as a single-layer carbon film on top of another surface, a technique known as chemical vapour deposition. This isn’t new either: ‘monolayer graphite films’, or graphene, were made this way in the 1970s. Typically, a mixture of methane and hydrogen gas is passed across a sheet of copper in a furnace at 800–1000°C, and a single carbon layer forms atop the copper; chemical processing can remove the copper

GROWING GRAPHENE FILMS

Researchers make large (centimetre-scale) graphene films by depositing carbon atoms from a vapour onto a copper surface. Roll-to-roll processing then transfers the graphene film from copper to another substrate.



Copper foil is heated to 800–1000°C in a furnace, where hydrogen (H₂) and methane gas (CH₄) are piped through. Carbon atoms grow as graphene film on top of the copper.

Adhesive 'thermal release' tape is attached on top of the graphene, by applying pressure between heavy rollers.

The copper is eaten away in a plastic bath filled with etchant (such as ammonium persulfate, a bleaching agent and oxidant).

On heating to 90–120°C, the adhesive tape unsticks, leaving the graphene clinging to a target substrate.

and deposit the graphene on a more useful substrate material, such as silicon dioxide (see 'Growing Graphene Films'). Tour says a laboratory using a US\$10,000 setup can make graphene sheets a few square centimetres in size. And while methane is the standard reactant, any carbon source will do: Tour has even put cockroach legs on top of copper and produced a thin graphene film.

A more sophisticated version of this approach yields graphene on an impressive scale. In 2010, researchers at Sungkyunkwan University in Suwon, South Korea, made transparent graphene films up to 76 cm across by depositing carbon atoms onto copper and etching the copper away². Electronics giant Samsung wants to use a reel-to-reel version of this technique to produce graphene film on a continuous roll — although Seungmin Cho, principal research engineer at subsidiary Samsung Techwin in Gyeongnam, South Korea, says for now the company is concentrating on making high-quality single sheets.

Materials scientists are working on ways to lift the graphene off the copper plate and onto insulating substrates — which won't interfere with electronic properties — without wrinkling the film or otherwise compromising its quality. Top of their list is silicon dioxide, the ubiquitous insulator used in silicon chips. In 2011, Tour showed that when nickel is placed atop silicon oxide, a layer of carbon film grows at each surface of the interface between the two; etching away the nickel leaves a bilayer of graphene coated directly on silicon oxide³. But the growth process takes place at 1000°C: temperatures high enough to wreck any dopants or pre-structured devices (such as transistors) in the underlying silicon. Tour is confident that graphene can be grown easily on insulating substrates. "I'm reviewing papers all the time that promise improvements," he says. One paper this year, by Soon-Yong Kwon and colleagues at the Ulsan National Institute of

Science and Technology in South Korea, promises a similar process that works at just above room temperature⁴.

Even a graphene film grown on a surface is far from perfect. The film forms when micrometre-sized patches of hexagonally-arrayed carbon atoms — each cluster a tiny fragment of graphene — sprout simultaneously from various points on the substrate. The clusters never stitch together perfectly, so the final graphene sheet looks like a patchwork quilt of carbon, with kinks of 5- and 7-membered rings where each patch joins with the next. The largest single perfect graphene patch ever created is reported by nanotechnologist Rodney Ruoff at The University of Texas at Austin, together with scientists at Texas Instruments in Dallas, who grew individual clusters up to 0.5 millimetres on a side⁵.

FROM SHEETS TO RIBBONS

Graphene film conducts electrons so well, it is hard to block their flow — a problem if one wants to use graphene in a transistor to control electrical current. But thin ribbons of graphene can be semiconducting, because the narrow channels alter the energy states of electrons. The problem is in how to make such a ribbon. Using lithography to slice up large sheets into graphene strips is, says Tour, "like using a chainsaw to do the finishing work in your home" — the ribbon's edges produced by this crude approach lack the uniformity required for electronic control.

One approach is to start with carbon nanotubes, which are essentially rolled up graphene. Researchers have variously unzipped them into flat sheets using metal catalysts, or etched 10–20 nm wide ribbons from them using ionized argon gas. But control of the edges remains a problem: only one particular geometry of the ribbon's edge will produce semiconducting graphene.

Physicist Walter de Heer at the Georgia

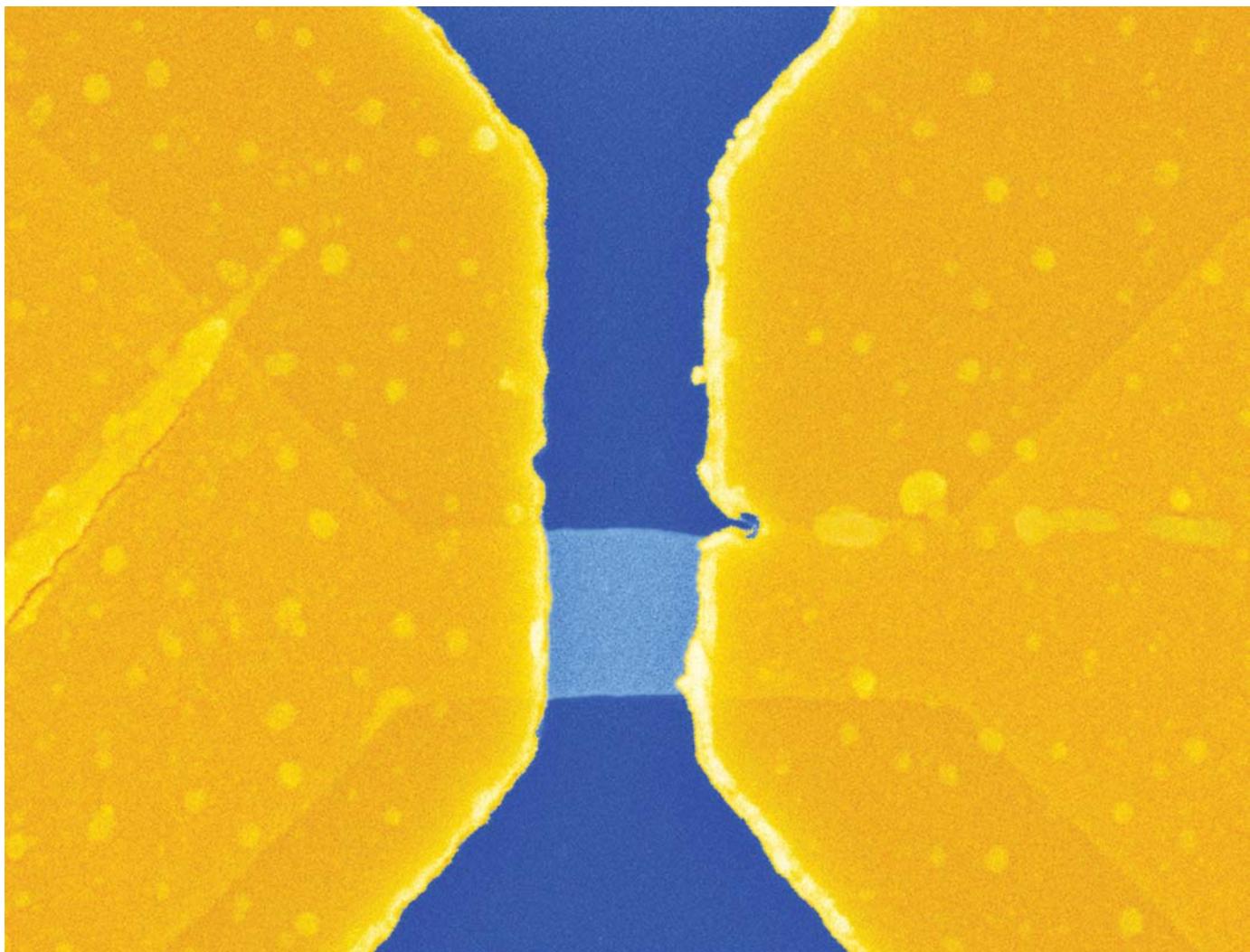
Institute of Technology in Atlanta takes a different tack: growing graphene ribbons directly in place on a substrate. He heats silicon carbide — a compound of silicon and carbon that has been mass-produced for over a century — to 1,600°C, and the silicon atoms on top escape, leaving carbon behind. To grow graphene ribbons exactly as required, de Heer notches a step into the silicon carbide; the graphene grows preferentially on the sidewall⁶.

The best graphene ribbons developed so far are those grown molecule by molecule. Chemist Klaus Müllen's team at the Max-Planck Institute for Polymer Research in Mainz, Germany, make flawless ribbons via chemical reactions between molecules based on hexagonal rings — each molecule chosen to produce an atomically precise graphene lattice when they fuse in long chains⁷. Müllen says his team can now make perfect graphene ribbons. The catch: they are only a few hundred nanometres long. De Heer envisages using similar chemistry to grow thin semi-conducting electronic interconnects between larger metallic areas of graphene — although such architectures and their applications are decades from fruition.

From the heights of molecular perfection, to the large, imperfect platelets on the graphene market: there's space for all production methods — though they seem worlds apart. ■

Richard Van Noorden is an assistant news editor at Nature.

- Xu, Z. & Gao, C. *Nature Communications* **2**, 571 (2011).
- Bae, S. et al. *Nature Nanotech.* **5**, 574–578 (2010).
- Zhiwei Peng et al. *ACS Nano* **5**, 8241–8247 (2011).
- Kwak, J. et al. *Nature Comms* **3**, 645 (2012).
- Li, X. et al. *J. Am. Chem. Soc.* **133**, 2816–2819 (2011).
- Sprinkle, M. et al. *Nature Nanotech.* **5**, 727–731 (2010).
- Cai, J. et al. *Nature* **466**, 470–473 (2010).



ANDRE GEIM, KOSTYA NOVOSELOV/SCIENCE PHOTO LIBRARY

Coloured scanning electron micrograph of a transistor composed of a graphene wire (centre), gold electrodes (dark yellow) and silicon (blue).

ELECTRONICS

Back to analogue

Trying to shoehorn graphene into a digital circuit isn't working. But there may be another potential path to glory.

BY KATHERINE BOURZAC

In July 2008, Tomás Palacios was sitting in an airport in Boston, Massachusetts, waiting for a delayed flight and thinking about his latest batch of graphene transistors. Yet again, they weren't working as they should. Palacios was excited when he learned about graphene only four years before, and he hoped this form of carbon could hasten a future generation of computer chips once silicon was pushed to its physical limits. But a major hurdle blocks the way: there seems to be no good way to make graphene transistors that can perform the on-off digital logic that

is the basis of computation. For the past few years, many researchers have been tinkering with graphene to make it behave more like a conventional material. Like Palacios, many had been disappointed. "It is very difficult to fight against nature," he says.

With hours to kill at the airport, Palacios started to mull over his results. Studying some of his latest current-versus-voltage graphs, he realized that the electrical behaviour of a single graphene transistor looks like that of a circuit comprising multiple silicon ones. He called one of his students and asked him to test a single graphene transistor against the silicon circuit. When he finally got back to his lab at

the Massachusetts Institute of Technology in Cambridge, he saw that the graphene transistor outperformed the silicon grouping — and that his group had just created a new branch of research.

NANOSPAGHETTI

Ever since Andre Geim and Konstantin Novoselov of the University of Manchester, UK, published their seminal *Science* paper in 2004 describing graphene's electrical properties, researchers have been trying to figure out what to do with the material. Some of this effort has been like trying to push a square peg into a round hole. "If you try to fit graphene into a

silicon box, it's a bit awkward," says James Tour, a chemist at Rice University in Houston, Texas. Now, many people are testing graphene in transistors to see what it does well, which might not be digital computing. They're shifting their focus towards speedier flexible electronics, more compact and efficient circuits that operate in unusual ways, and high-performance analogue electronics that could improve telecommunications and mobile data streaming.

The faster that electrons can move from one end of a transistor material to the other — that is, the higher its mobility — the faster that information can move around a computer chip. When Geim and Novoselov wired up flakes of graphene, they found that electrons sailed swiftly through them. Indeed, electrons in graphene behave as if they have no mass. Graphene's mobility at room temperature can be as high as 100,000 square centimetres per volt per second — some 70 times higher than it is for silicon.

In addition to speed, graphene offers a potential size advantage. Consisting of a single layer of carbon atoms, the material is effectively two-dimensional. This means it should be possible to make computer switches even smaller than the silicon transistors that form the backbone of the US\$250-billion semiconductor market. "We would like to find a new switch, a material to replace silicon," says Luigi Colombo, a researcher at Texas Instruments in Dallas.

Graphene might also be easy to manufacture. Fengnian Xia, a researcher at the IBM Watson Research Center in Yorktown Heights, New York, says his lab can already grow sheets of the material on the silicon wafers that the semiconductor industry uses for large-scale chip manufacturing. Compatibility with existing infrastructure is a big plus of graphene over other new materials. Carbon nanotubes, for example, have similar properties to graphene but fail the infrastructure test: they are difficult to align on a surface, tending to tangle up like spaghetti.

MIND THE GAP

The big problem with making digital switches from graphene is that the material lacks an electrical property called a bandgap — essential for digital logic. To understand this phenomenon, it's helpful to look at a conventional silicon transistor, the type found in the hundreds of millions of today's computer chips. These transistors have four main parts, a silicon channel and three metal terminals: gate, source and drain. An electric voltage applied at the gate controls the flow of current from the source electrode through the silicon channel to the drain electrode. Silicon has certain energy states where it's conductive, and others where it's not.

The energy difference between those two states is called the bandgap. By applying particular voltages to the channel

through the gate, it's possible to switch silicon's energy level between the conductive state, to turn the transistor on and make a 1, and the insulating state, to turn the transistor off and make a 0. Getting that zero is critical for efficient digital logic, and to keep power consumption in check. Because there are hundreds of millions of transistors in a microprocessor, it's important to be able to turn them off when they're not in use, otherwise their power requirements would be unsustainable. But graphene does not naturally have a bandgap and can't make a zero. Researchers have developed ways to overcome this problem. Graphene has been doped with chemicals, punched with tiny holes, and exposed to all kinds of abuse to induce a bandgap.

Unfortunately, such interventions also compromise graphene's charge mobility, so the resulting transistors are not fast enough to be worth the bother. "Nobody has made a switch out of graphene that is technologically feasible

and comparable to silicon," says Walter de Heer, a physicist at the Georgia Institute of Technology in Atlanta, who made some of the first graphene transistors. Novoselov agrees, but notes that people are

still coming to grips with graphene. "We knew from the beginning that it would be challenging to make graphene work for logic applications," he says.

So researchers are now using graphene transistors in a field that plays better to its strengths: analogue electronics. Analogue devices, the electronics at work in communications equipment such as receivers, transmitters and amplifiers, don't speak in zeros and ones but operate over a range of voltages. Wherever an electronic device meets the outside world, there are analogue electronics. The chip board in a smartphone has many. Besides processing audio, it handles photographs taken with the camera, while other circuits stream video. So although much of what we think of as information processing now occurs in the digital realm, "analogue sets the speed limit on data", says electrical engineer Deji Akinwande, who studies carbon nanomaterials at the University of Texas at Austin.

SPEEDING COMMUNICATIONS

Phaedon Avouris, a materials scientist at IBM, has been working on graphene transistors since 2006. Working independently, his group and Philip Kim's at Columbia University in New York made the first graphene transistors with a bandgap by slicing the material into ribbons tens of atoms wide with the help of an electron beam. The procedure is complicated, and the resulting transistors are slow. "It became clear that this will not go too far in a practical application," he says. "If graphene is really the

miracle material, use it as it is and don't try to change it into something else."

For Avouris, using graphene properly means taking advantage of its high mobility in high-frequency analogue electronics and its ability to integrate well with other materials. Generally, the faster the analogue electronics can switch — the higher their frequency — the less noisy, or error-prone, they are. Higher-frequency electronics can produce stronger signals that can travel farther, and they use less energy than low-frequency electronics. And signals sent at higher frequencies have lower interference, which means more of them can be nestled together in a given band of the spectrum for faster data transmission.

The analogue electronics in consumer goods are, like digital electronics, based on silicon. For niche applications, such as encoding military messages, and transmitting and receiving signals in high-power mobile-phone base stations, speedier electronics made from more expensive materials are required. These so-called 'III-V' materials (named after those two groups in the periodic table their elements come from) include gallium arsenide and indium phosphide. Some III-V transistors operate at frequencies as high as 1 terahertz (1,000 gigahertz). But whereas III-V materials aren't compatible with silicon electronics and need to be placed on their own chips, graphene is compatible with silicon.

So far, no graphene circuits have reached terahertz speeds. In 2010, Avouris made transistors with speeds of 100 GHz — about ten times as fast as the best silicon transistors. Avouris believes the only way forward is to use the equipment already on the floor of today's semiconductor factories, so his group has been mostly using graphene films on the surface of the same wafers used to make computer chips. The IBM result has been overtaken by Xiangfeng Duan, a chemist at the University of California, Los Angeles, with a 300-GHz graphene transistor made with graphene flakes. But these flakes are not as manufacturing-friendly as the films of graphene that the IBM group used. Both groups achieved these speeds by taking good care not to damage the graphene — a tough challenge because other materials must be laid on the graphene in order to make a transistor.

Electronics require integrated circuits rather than single transistors. During the summer of 2011, Avouris' team made the first graphene integrated circuit on a wafer, reaching a record speed of about 10 GHz. The circuit combined a graphene transistor with two inductors to make one of the workhorse devices of telecommunications: a frequency mixer. Frequency mixers combine a signal to be transmitted — say, the sound of a voice as picked up by a mobile-phone microphone — with a reference signal so that it can travel through the air. Frequency mixers also unscramble such signals at the other end. Similar efforts to develop

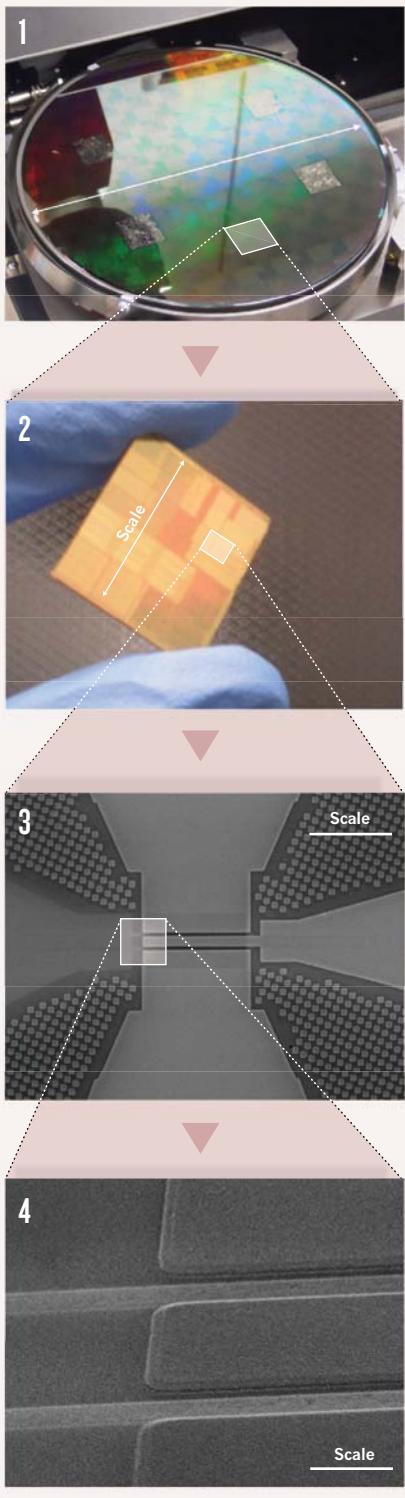
ON NATURE.COM

For some of the latest research on graphene transistors: go.nature.com/2d2ovl

READY TO GO

One of the advantages of graphene electronics is that they can be manufactured using equipment and processes already used by the chip industry.

1. The 8-inch wafer is covered in graphene transistors.
2. The wafer can be cut into chips.
3. Microscope zooms in on a single transistor.
4. View of the gates (horizontal bars) that control the transistor.



graphene-based analogue electronics are underway at HRL Laboratories in Malibu, California, and at several universities.

Many challenges remain. The IBM frequency mixers won't be found in mobile phones any time soon because the signals they produce are too weak, a problem that originates at the transistor level and that Avouris and others are trying to address. The hundreds of gigahertz speeds reported for graphene devices often ignore an important fact, says Columbia University electrical engineer Ken Shepard. The faster these devices operate, the weaker the signal they generate. When the signal gets overwhelmed by noise, the circuit becomes useless. These effects, dubbed parasitics, mean that in reality, graphene transistors are not much faster than silicon ones. "For these transistors to be earth-changing they have to be at least an order of magnitude better than silicon and that's just not there — they are a factor of two better," Shepard says. Whether these problems are fundamental or are engineering challenges isn't yet clear, he says. "As more designers look at this material, the technology issues will fall away, and we will get a better idea of what the intrinsic limitations are."

EMBRACING ECCENTRICITY

Rather than try to engineer graphene for high-frequency devices that might one day supplant silicon-based electronics, some researchers are trying to beat new paths. Over the past couple of years, says MIT's Palacios, "people are focusing more and more on applications for graphene that silicon cannot do".

Graphene is stretchy, flexible, transparent and strong; silicon is none of these things. Because graphene has other attractive properties, the material doesn't have to beat silicon to be useful for digital or analogue electronics, says the University of Texas's Akinwande. He believes graphene is suited for applications such as ubiquitous computing or flexible, transparent portable electronics that don't break when dropped. Here the material is competing with very-low-electron-mobility materials such as semiconducting polymers. "The goal of flexible electronics is to make them waterproof, shockproof and environmentally friendly," says Akinwande.

Akinwande is working on making graphene circuits on plastic. The challenges are in the manufacturing, he says. Compared to the smooth surface of a silicon wafer, plastic is bumpy and graphene is extraordinarily sensitive to tiny flaws. Akinwande is working to establish partnerships with companies, including 3M, a technology company based in St Paul, Minnesota, to get plastic that's suitable for graphene transistors. So far, graphene-on-plastic transistors have an intrinsic speed of 5 GHz — significantly slower than Duan's 300-GHz transistors, but faster than the best organic polymer transistors, which share graphene's flexibility.

"Graphene will go first into applications

that only graphene can serve," says Palacios. He also sees promise for graphene in flexible electronics. And from that day in the airport in 2008, he's been working on turning his frustration with graphene's lack of a bandgap into an advantage in highly efficient graphene devices: single transistors that do the work of many silicon devices, and do it better. A single graphene device could do the work of "20 silicon transistors", Palacios says, which would "reduce power consumption and save space".

Graphene transistors can multitask because as the voltage varies, they reach a minimum current and then switch back and forth between carrying negative and positive charges. Kartik Mohanram, an engineer at the University of Pittsburgh in Pennsylvania, likens such ambipolarity to an eccentric water tap. It would be as if when you turn off the cold tap, the flow tapers off before releasing a gush of hot water.

Most engineers avoid graphene's ambipolarity, making sure these devices operate only in voltage ranges that conduct either positive or negative charges, but not both. By varying the voltage on a graphene transistor all along that scale from positive to negative, engineers are able to get a device that performs more functions than a conventional transistor. Palacios says that one of the ambipolar transistors he designed can convert a digital signal into a Bluetooth signal that can be sent between, for example, a mobile phone and a wireless headset. The conversion requires an analogue device called a phase-shift-key modulator. When made in silicon, this kind of modulator requires an integrated circuit chip. A single ambipolar graphene transistor, Palacios has shown, can do the job all by itself.

Mohanram admits that the idea that ambipolarity can have practical uses is not yet widely accepted by electrical engineers.

"The big challenge is whether we can do this on a large scale," he says. While circuit designers can come up with a new idea after a week's thought, it can

take a year to make and test just a few devices. Engineers are still learning how to work with graphene, and new ideas come along faster than they can be tested.

Ultimately, for graphene to make a big impact in electronics, the material will have to live up to its hype, and then some. As Texas Instruments' Colombo puts it, silicon "is a very tough act to follow". But patience could be the watchword. After all, it was only in 2004 that a large number of researchers began working with the material. "Maybe graphene will revolutionize electronics," says de Heer, the Georgia Tech physicist. "But not tomorrow." ■

Katherine Bourzac is a science journalist based in San Francisco, California.

The bionic material

Graphene could make an ideal basis for a medical repair kit.

BY CHARLES SCHMIDT

More and more people are having their ruined body parts replaced with prostheses interconnected to the nervous system. Advances in graphene technology might bring these artificial devices to their senses — in bionic eyes and ears.

Graphene is impervious to the harsh ionic solutions found in the human body. Moreover, graphene's ability to conduct electrical signals means it can interface with neurons and other cells that communicate by nerve impulse, or action potential. These features have made graphene a material of some promise in next-generation bionic technology.

In November 2011, Jose Garrido, a nanotechnologist at the Walter Schottky Institute in Munich, Germany, took a big bionic step when he showed that arrays of transistors made of graphene can detect action potentials in heart cells.

Garrido's work was a welcome breakthrough, says Philippe Bergonzo of CEA- LIST in Saclay, France; Bergonzo is coordinator of NeuroCare, a European project looking into using carbon-based implants in ears, eyes and the brain. Apart from its stability and favourable electronic properties, graphene is also flexible, so it can be wrapped around delicate tissues. No other material shares all these features, Bergonzo says, adding that graphene opens up research opportunities in neural prosthetics.

To make graphene transistors, Garrido uses a method called chemical vapour deposition (see 'Beyond sticky tape', page S32) and grows a layer of cardiomyocytes directly on top of the array.

HOW IT WORKS

Like other electrogenic cells, cardiomyocytes use action potentials to pass electrical signals along from cell to cell. Each of those voltage spikes changes the local electrostatic environment by inducing a flow of ions in the channel in the electrolyte separating the cells and the transistors. Garrido's transistors respond to this ion current by altering their electrical resistance, explains James Hone, who specializes in nanoscale devices at Columbia University in New York. The fluctuation in resistance constitutes a detectable signal between cardiomyocytes.

Unlike silicon transistors, those made of graphene can't be switched off — its physical properties don't allow for that, Hone says

— and this means they're not suited to digital applications where devices must be able to generate ones and zeroes (see 'Back to analogue', page S34). What graphene transistors are good at, according to Garrido, is biological sensing — the sort of task that eyes and ears perform. In such analogue applications, the ability to switch off is not critical, and graphene's distinct qualities come to the fore. Should silicon transistors be used in the human body, they would need to be coated with metal oxide to boost their stability in solution, Garrido says. Those

expected budget of US\$6.3 million — the optimal neural device will both sense and stimulate cell activity. "That's where we expect a breakthrough from graphene," Bergonzo says. "We'd like to know how neural networks are behaving in real time so we can stimulate them more effectively. Otherwise, you risk giving too much stimulation, or not enough."

Serge Picaud, principal investigator with the Vision Institute in Paris, involved in the NeuroCare project, adds that because graphene is so thin, it could improve the interface between retinal implants and eye tissues. And those closer connections, he says, could improve sight.

REMAINING CHALLENGES

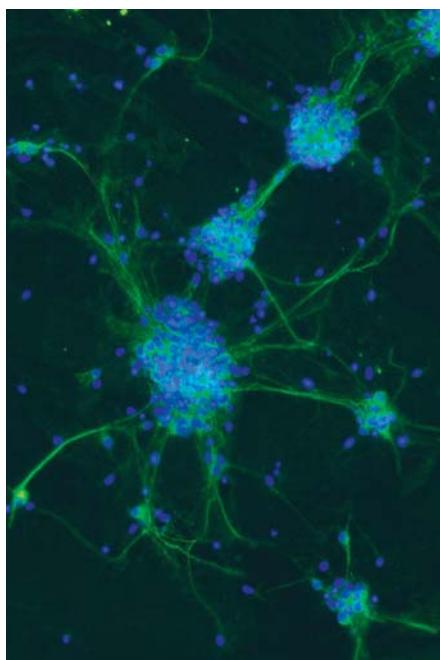
Hone emphasizes that research into graphene-based bioelectronics is in its infancy. Scientists still face fundamental challenges in manufacturing, he says. Chemical vapour deposition, in particular, doesn't generate perfect graphene, and this limits the material's electronic performance (see 'Beyond sticky tape', page S32).

Moreover, materials scientist John Rogers at the University of Illinois in Urbana-Champaign cautions that silicon is still a contender. Silicon can be fabricated into structures as thin as 10 nm; while that doesn't match the 1-nm dimensions possible with graphene, it might just work. What's more, Rogers points to a deep scientific and engineering base for silicon in the semiconductor industry. Rogers says that researchers are finding new ways to encapsulate implanted silicon devices so they don't harm tissue. Still, Rogers sees tremendous opportunities for graphene in bioelectronic sensing because of the carbon material's much lower electrical noise. He envisions a hybrid approach that takes advantage of the strengths of both silicon and graphene. "I don't think you'd want to make your entire electronic system out of graphene," he says. "You could use ultra-thin silicon for switching and processing hardware and then use graphene to establish your interface with tissue."

For his part, Garrido says he's now working to stack graphene transistor arrays on flexible substrates, such as the biocompatible polymers parylene and polyimide, each mechanically and heat stable. Graphene is clearly an "enabling material", he says. "And it has properties that could lead to a whole new generation of neural devices." ■

Charles Schmidt is a science writer based in Portland, Maine.

1. Hess, L. H. *et al.* *Adv. Mat.* **23**, 5045–5049 (2011).



Cortical neurons being grown on graphene for use as biological implants (nuclei are stained blue).

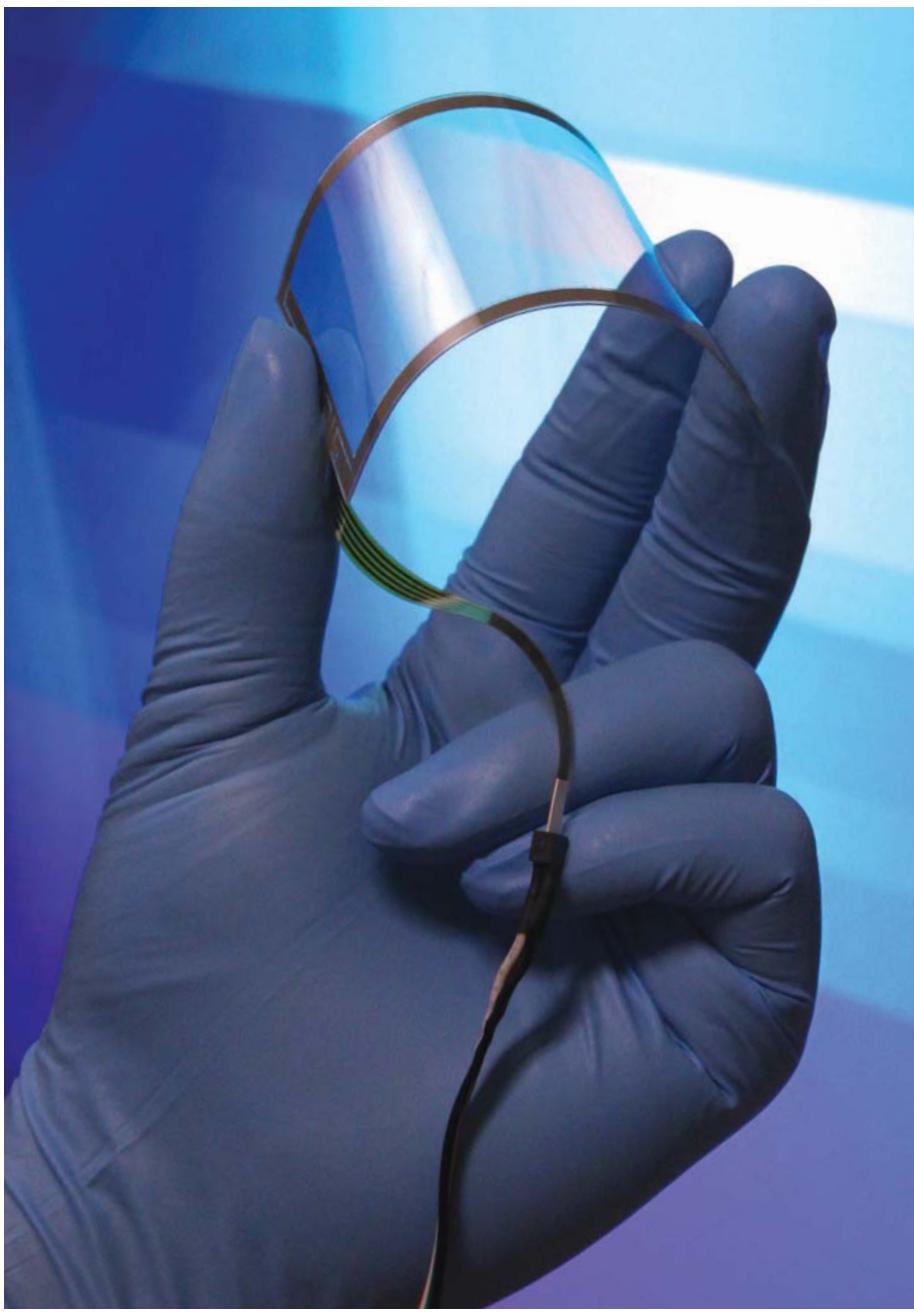
layers trap ions that produce noisy interference and thus degrade signal quality. Graphene transistors, on the other hand, don't require an oxide coat so they generate less intrinsic noise, which enables them to detect the faint signals (generally below a few hundred microvolts) of cell communication.

Graphene is not the only carbon-based material with bionic potential. Diamond nanocrystals show promise in retinal implants to treat blindness, says Bergonzo. But diamond is solid,

inflexible and a poor conductor. For Bergonzo and his collaborators at NeuroCare — which is expected set to launch in March 2012 with an

NATURE.COM

For some of the latest advances in prosthetics: go.nature.com/hxbyrm



Touch screens with a graphene electrode could bring flexibility to mobile phones and other devices.

OPTOELECTRONICS

Come into the light

Transparency across the spectrum combined with electronic prowess makes graphene an ideal photonic material.

BY NEIL SAVAGE

Graphene's first major appearances outside laboratories may be at the intersection of electronics and optics, in the form of transparent electrodes. Such electrodes — which carry current but allow light to pass through — let touch screens operate, collect current from solar cells without blocking the light needed to produce that current, and inject current to power organic light-emitting diodes (OLEDs) while allowing most of the photons to escape.

Much graphene research has focused on the material's remarkable electronic abilities and the possibility of using it to build transistors. However, it competes there with a well-established technology — silicon. As James Tour, a chemist at Rice University in Houston, Texas, puts it: "Silicon has had a 50-year head start, trillions of dollars and millions of person years." So some researchers are starting to explore possibilities that combine graphene's electronic properties with its light-handling characteristics, as well as other attractive properties such as strength and flexibility. Transparent electrodes today generally rely on indium tin oxide (ITO). But indium is rare, and its price is rising as demand for smartphones and tablets soars. And because ITO is brittle, it must be deposited on an unbending surface (usually glass), thus placing constraints on device design. With a thin film of graphene as a conductive material, manufacturers could build flexible organic solar cells; OLED panels could be wrapped around furniture, automobile frames or a wrist. Graphene electrodes could make possible lightweight smartphones that could be folded or rolled up and whose screen won't crack when dropped. At the Consumer Electronics Show in early 2011, Korean electronics firm Samsung displayed a flexible active-matrix OLED phone screen based on graphene, but the company has not announced any products incorporating this technology.

Tour concedes that graphene's performance characteristics are not as good as ITO's, at least not yet. A thin film of ITO lets about 90% of the light pass through it, and has electrical resistance of 50 ohms. A single-layer sheet of pristine graphene allows more light through — 97.7% — but has far higher resistance, between 2,000 and 5,000 ohms. Adding more layers of graphene improves resistance but worsens transparency; four layers provide transparency roughly comparable to ITO's but, at 350 ohms, it still has seven times the resistance.

But graphene may not need to be as good electrically as ITO to get into the marketplace — being cheaper may be enough. And in the meantime, researchers are working to enhance the carbon monolayer's performance. Tour is building a touchscreen that contains a thin mesh of aluminium nanowires between the substrate and the graphene. Only 5 micrometres in diameter, the wires are too slender to block light but they make the device dramatically more conductive. A screen with 90% transparency now

has a resistance of only 20 ohms. Researchers at Samsung and at Sungkyunkwan University in Seoul have collaborated to obtain similar results — 90% transparency with 30 ohms resistance — by mixing in nitric acid. And these South Korean groups have shown they can produce sheets of the improved graphene that measure 30 inches diagonally, using a roll-to-roll process that would be convenient for manufacturing.

Finding efficient production processes is part of the challenge of introducing graphene transparent electrodes. Tour has devised a method to create thin ribbons of graphene by treating carbon nanotubes with sulphuric acid, causing them to unzip into ribbons. He then attaches to the ribbons' edges molecular groups that allow the material to be suspended in a solution and sprayed onto a surface. This approach could allow transparent electrodes to be painted onto large-area solar cells or OLEDs, potentially providing an inexpensive alternative to silicon solar cells.

Andrea Ferrari, a nanotechnologist at the University of Cambridge in the United Kingdom, has made graphene-based inks that an inkjet printer can use to make transparent electrodes and circuits; graphene ink and polymers could be used to create thin-film solar cells, he suggests. These wouldn't convert as much light to electricity as a silicon solar cell, but because they're so lightweight and transparent they could be installed on the facades of buildings, capturing sunlight that would otherwise go to waste without inducing any structural strain or ruining the aesthetics. A thin-film solar cell could be transparent enough to cover a window pane, yet still produce enough energy to make a dent in an electricity bill. Filling niches where silicon isn't appropriate could be a small but still valuable role for graphene, Ferrari says. "Even though it works much worse than the state of the art, it has the advantage of being flexible and cheap."

GAPS IN THE SPECTRUM

Graphene's most attractive optical property, according to Ferrari, is that it "works at absolutely any wavelength whatsoever". That is, the material absorbs the same fraction of light — roughly 2.3% — all across the spectrum, from the ultraviolet into the far infrared. This property arises from the electronic structure of graphene: it lacks a bandgap — a range of energy states where electrons are prohibited from existing. Graphene's ability to absorb across the spectrum is important because the way light interacts with different materials — whether it's absorbed or scattered or passes through — varies depending on its wavelength. One wavelength may be ideal for cutting a certain material, another for detecting particular organic molecules. But at some wavelengths, there's no device to perform those tasks, at least one that isn't prohibitively expensive. Much of the electromagnetic spectrum, Ferrari notes, has "never been exploited, because it's hard to make semiconductors that work at those wavelengths."

One such untapped portion of the spectrum is between the far infrared and microwaves, or

frequencies between about 100 and 10,000 gigahertz (0.1–10 terahertz). Because such terahertz radiation penetrates a wide variety of materials but is non-ionizing, it has potential applications in security scans, environmental sensing, and medical imaging. Currently, however, sources and detectors that operate at those frequencies are few and complex; some terahertz detectors require cryogenic cooling, whereas a graphene detector could work at room temperature. "Some people refer to it as the last gap in the electromagnetic wavelengths," says physicist Feng Wang, head of the Ultrafast Nano-Optics Group at the University of California, Berkeley.

Wang built a terahertz detector based on nanoribbons of graphene ranging in width between 1 and 4 micrometres. The detector is based on plasmon resonance: when light strikes a material whose electrons resonate with a frequency that matches the light's wavelength, the result is oscillations in electron density called plasmons, which can enhance the strength of a signal. (The colours in stained glass windows are so vibrant because of plasmon resonance between visible light and metal nanoparticles embedded in the glass.) In graphene, this plasmon resonance happens at terahertz frequencies. Wang designs devices to detect particular frequencies by choosing the width of the ribbons, and then further tunes the device by applying an electrical field. Creating such a detector is the

Although the terahertz range is the area that's most sorely lacking emitters and detectors, graphene could fill in other gaps in the spectrum as well, because "you have a sort of universal detector", says Phaedon Avouris, head of the nanoscale science and technology group at IBM's Watson Research Center in Yorktown Heights, New York. Current technology does not offer fast photodetectors at mid- to far-infrared wavelengths; Avouris's team built an experimental graphene-based detector for this region of the spectrum that operates at a speed of 40 GHz. And, Avouris contends, because of imperfect measurement tools, it's possible that the graphene device could actually operate at 500 GHz or higher.

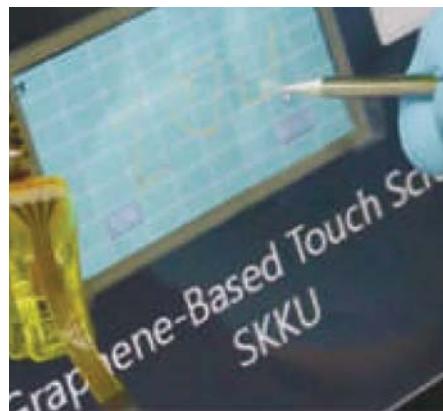
As for emitters, Ferrari says graphene might soon be used in ultrafast lasers. Such lasers produce short, high-energy bursts of light, lasting only picoseconds (trillionths of a second) or femtoseconds (thousandths of a picosecond). Delivering a lot of energy in a short burst lets the lasers cut very precisely without damaging nearby material or tissue; they're used in applications from eye surgery to printed-circuit-board processing. Most ultrafast lasers use a technique called mode-locking, in which a device known as a saturable absorber soaks up continuous laser emission and releases it in a series of pulses. The absorbers are usually semiconductor mirrors. But such systems are complex to build, and they generally emit light at only a narrow range of wavelengths. Using graphene as the saturable absorber would not only allow a simpler design but also permit broader spectrum operation.

SMALL AND FAST

In addition to emitting and detecting light, photonics systems often require an optical modulator, which encodes signals onto light beams. Optical telecommunications use a modulator built of lithium niobate, and that's unlikely to be displaced for the majority of applications, says Wang. His group demonstrated a graphene modulator about 25 square micrometres in size — 4 million times smaller than a conventional modulator made from lithium niobate. Such a small device could be integrated on a computer chip for optical interconnections, replacing copper wires with light beams and clearing a bottleneck for processing speed. The modulators should also work up to 10 times faster than existing devices and, unlike lithium niobate, could be used in the mid-infrared, giving system designers more flexibility.

It will take some time to figure out the impact on photonics of these emerging graphene technologies, Ferrari says. "Graphene optoelectronics is only three years old," he says. "We need to give it another two or three years to see where we are going." By that time, this toddler technology might be starting to light the way forward in ways not yet conceived. ■

Neil Savage is a science and technology journalist based in Lowell, Massachusetts.



A graphene-based touch screen in operation.

beginning of developing a set of tools for the terahertz range, he says.

Graphene may be valuable not only as a detector of terahertz radiation, but also as a source, Wang says. There are already quantum cascade lasers that emit at frequencies of 10 to 100 THz. The most useful frequencies, however, lie between about 500 GHz and 4 THz, and the only existing sources in that range are limited to very low output power. Wang points out that as researchers push to create faster transistors that operate at these speeds, they could in effect be creating an emission source in that range as well — just as conventional megahertz-speed transistors produce radio waves. Pushing emitters to higher and higher frequencies opens up the possibility of higher bandwidth communications.



Q&A Tomás Palacios Taking charge

Nature Outlook talks to the first director of the MIT's Centre for Graphene Devices and Systems, which was created in July 2011 to foster collaboration among academic, industrial and government groups studying this form of carbon.

What is the appeal of graphene?

Graphene is an extreme material. There is nothing thinner or with better transport properties than this one-atom-thick layer of carbon atoms. These and other properties offer tremendous opportunities in electronics, materials science and chemical engineering. In addition, the initial work on graphene has motivated research in many other layered materials, such as boron nitride and tungsten disulphide. All these materials offer the ultimate nanotechnology dream: atomic control at the macroscale.

What's the big deal about single layer?

High-frequency transistors require the charge carriers to be as close as possible to the gate electrode that controls the flow of current. In conventional materials, like the silicon used in ordinary transistors, the electrons spread in the vertical direction by at least 5 to 10 nanometres. This reduces the gate electrode's control of these electrons, which in turn limits how small and fast the transistors are. Confining the carriers to one layer of atoms preserves that control. In semi-transparent electrodes for flat panel displays, graphene's single-atom-thickness minimizes light absorption and makes for almost transparent layers.

How well characterized are graphene's properties?

We understand its low-energy transport properties quite well. However, we still need a lot more work to develop devices and systems that can benefit from these properties. One example is the very long mean free path of charge carriers in graphene. In many cases, charges can travel across graphene for more than one micrometre without bumping into anything. That is orders of magnitude farther than in other materials. This reduced scattering could enable a new generation of devices inspired by the old vacuum-tube electronics, where the electrons are controlled not only by electric fields but also by magnetic fields. Somewhat related is the possibility of measuring the quantum Hall effect at room temperature in this device — a phenomenon resulting from the interaction between an electric current and a magnetic field. That could be useful when you need a precise measure of resistance or voltage — such as those needed in metrology or high-speed communications.

How widespread are the potential applications of graphene in electronics?

The first commercial application will probably be as a semi-transparent electrode in solar cells, light-emitting diodes (LEDs) or touch screens. The metal used in transparent conducting films is expensive; graphene could be much cheaper, and much more flexible and robust. Several companies, including South Korea's Samsung and Spain's Graphenea Nanomaterials, either already have some initial products or plan to have

prototypes and products on the market soon.

Another interesting application is in microprocessors. As transistors get smaller and smaller, the diameter of the wires connecting them also needs to shrink. With conventional metallic conductors, defects and grain boundaries in the structure of crystalline metals can cause electrical resistivity to increase quickly at scales smaller than 50 nm. Because graphene is a single crystal, there are no grain boundaries, and its resistivity should stay roughly constant with diameter. That makes it ideal for low-resistance interconnections in sub-20-nm transistors.

Once some of these initial applications are successful, many others will follow. Look at optics, for example. Graphene absorbs light uniformly across the entire spectrum, from the infrared to the ultraviolet — that's true of very few other materials. We are taking advantage of this to develop a new generation of night-vision devices, which will be cheaper and able to cover larger fields of view. We hope to build an advanced prototype in 2012.

Those are improvements on known technology. What about things that would be possible only with graphene?

The development of a room-temperature quantum Hall effect device could enable a new generation of electronic circuits based on the precise voltage and resistance references that it would enable. The high performance of graphene devices over very large areas would also allow low-cost electronics in novel forms; think of a coffee cup that displays today's news headlines, windows that display the outside temperature, and clothing that acts as antennae to extend the range of phones. Researchers are also investigating the use of graphene devices to steer electron beams in a way similar to what we do today in optics with light. This could revolutionize optical communications and enable new approaches for quantum computation and reconfigurable computers.

Where is graphene on the development continuum from research to manufacturing?

Devices involving semi-transparent electronics are very close to the market. Other applications, such as the use of graphene in computation or in very-large-area electronics, will take quite a bit longer. This is, among other reasons, due to the highly interdisciplinary work required for these applications. Methods to grow high-quality, large-area graphene films need to be developed; the fabrication technologies have to be modified to prevent the degradation of this one-atom-thick material during device fabrication; and, finally, new ways to integrate graphene devices and systems with more conventional electronics need to be invented.

What more futuristic applications do you imagine?

We are working on high-frequency graphene electronics — on clothes, paper and other

items — to communicate with smartphones. We are collaborating with industrial partners on graphene-based patches to monitor glucose levels in the blood of people with diabetes. Every object would just need a graphene-based sticker to connect to the Internet. In collaboration with the International Iberian Nanotechnology Laboratory in Braga, Portugal, we are developing a graphene sensor to detect *E. coli* in food.

Another promising area is the use of graphene in supercapacitors. The key here is to increase the surface-to-volume ratio of the electrodes that form the capacitor. To do this, researchers have recently used forests of carbon nanotubes, but only half of the nanotube's surface is exposed to the electrolyte; the other half is shielded by shape of the nanotube. With graphene, the entire surface — both sides of the layer of carbon — is available. This greater exposure enables higher energy storage. Something similar happens with conductive fibres. Carbon nanotubes are now used to make polymer fibres conductive, but the nanotubes only touch at a single point. When two graphene flakes touch, on the other hand, they do so over a much larger area. This more intimate contact increases the electrical and thermal conductivity.

Does any particular electronics stand to benefit from graphene?

Terahertz electronics presents a good opportunity. Most of the digital electronics in our computers work at a frequency of a few gigahertz (GHz). The analogue electronics used for wireless communications operates in the 2–40 GHz range, and a few radar and imaging systems work at 100–200 GHz. However, we have limited options today if we want electronics operating at frequencies in the 300 GHz to 3 terahertz (THz) range.

What does graphene offer to terahertz electronics?

Graphene's high charge mobility, in combination with its one-atom thickness, makes it attractive for filling the 'terahertz gap'. Devices operating at this range would work in advanced imaging systems, secure communications and chemical sensors based on terahertz spectroscopy. We need to find, however, a solution to graphene's lack of bandgap — a range of energies that the electrons are unable to occupy.

What specific areas of technology will the centre focus on?

We will soon see a pipeline of graphene applications beginning to emerge, and our centre will be an integral part of that. For that, we pursue a multidisciplinary approach where we are bringing together physicists, materials scientists, and chemical and electronic engineers to help move graphene from being a really amazing material to being an engineering reality that opens new industries. That requires collaboration between academia, industry and government.

What kind of support is the centre receiving from industry?

If we want graphene to be successful beyond basic science it is important that industry adopts it and develops products with it. In the centre, we have industrial members who provide highly complementary expertise and are helping to create a 'graphene ecosystem'. The Santa Clara, California, company Applied Materials, for example, a leader in processing equipment for the semiconductor industry, is excited about adapting its technologies to graphene. Another one of our member companies, STMicroelectronics, is working with us in the development of the next generation of graphene chemical and biological sensors, while the Dutch company DSM is contributing its expertise in chemistry and advanced materials.

Where does the government enter the centre's planning?

Agencies such as the Office of Naval Research and NASA have been supporting graphene research for many years through several programmes. We are currently working with them to develop detailed roadmaps for device and system insertion.

Does your centre plan to study and develop any two-dimensional materials other than graphene?

Graphene is an amazing material but it isn't necessarily the best material for every application. We have extensive work on 2D materials such as boron nitride or tungsten disulphide, materials that could be very interesting in high-power, high-temperature applications, among others.

Every technology faces hurdles in going from lab to market. What particular challenges does graphene face?

We face several. First, the commercialization of graphene at affordable prices requires the development of mass production techniques. Second, many applications, most notably digital electronics, require a material that has a bandgap — a demand that would seem to rule graphene out. Finally, we need to create an ecosystem for graphene in which companies interested in developing growth and processing equipment talk to companies interested in devices and materials.

What do you hope the centre will have achieved in its first five years?

Our ultimate goal is to transform graphene from a wonder material to an engineering reality. By 2016, I hope we will have generated commercial applications in electronics, chemistry and materials that take advantage of graphene's unique properties. I am certain that it will transform many fields. At the MIT Centre for Graphene Devices and Systems we are trying to speed up this process. ■

Interview by Peter Gwynne, a freelance science writer based in Sandwich, Massachusetts.

PERSPECTIVE



A means to an end

Exploring graphene's chemical properties reveals a world of potential away from the purely two-dimensional, says **Rodney Ruoff**.

Much of graphene's appeal comes from its extraordinary electronic properties that might one day make super-high-speed devices a reality, but graphene has more to offer than just its high electron mobility. It has other attributes deserving of attention — in particular, taking advantage of its chemistry to make new materials.

We recently made an activated microwave-expanded graphite oxide (a-MEGO) by activating a graphene-like precursor with potassium hydroxide¹. This porous material is composed of atom-thick walls of trivalently bonded carbon and pores 0.6–5.0 nanometres in diameter. It has a remarkably high surface area — 3,100 metres squared per gram — some 17% higher than anything possible with graphene itself. And it came as a complete surprise to us; its carbon lattice includes heptagons and octagons that form regions of so-called negative curvature — or the familiar saddle shape. We expected to get separate perforated thin sheets with many atoms along the sheet edges.

Saddle-shaped 'carbon nanofoams' have never been made in bulk and have mostly remained a theory; if they could be made in a continuous sheet form tens of micrometres thick, they would likely be the ultimate electrode material for very high energy density and ultrafast supercapacitors (thus having extremely high power density as well), which store electric charge. Until recently, supercapacitors have had much lower stored charge per unit weight or volume than batteries. But carbon a-MEGO gives supercapacitors energy densities comparable to that of conventional lead-acid batteries — but they are able to deliver electricity much faster. Supercapacitors with such high-performing carbons could find use in the electric grid, consumer electronics, power tools, and to replace the lead-acid batteries in cars, but with a much longer cycle life into the hundreds of thousands (thus surpassing the lifetime of most cars).

Treating graphene as a reactant rather than a product opens up other possibilities for novel and unexpected things. Graphene could be transformed into diamond without the conditions of the Earth's core. For example, stacks of graphene layers could be converted to a product that has a mixture of trivalently and tetravalently bonded carbon atoms — or perhaps only tetravalently bonded, like diamond. Such a material might have a tensile strength far exceeding the 5.6 gigapascals achieved by the carbon fibres that hold the world record in strength. There is rapid progress towards making metre-size sheets of graphene; on the research front, there is intense interest in improving reel-to-reel production to make, say, kilometre lengths of graphene

foils. It is time to consider clever ways to convert such graphene-based foils to nanometres-thick diamond foils. This can probably be done using low-energy chemistry, either by building up from monolayer graphene or by stitching adjacent layers together². This route might provide a form of diamond that would be ultrastrong over lengths of metres or longer, while keeping diamond's alluring properties intact. Doping very large-area diamond foils with nitrogen or other elements could render it electrically conductive for nanoelectronics — a product that might use nitrogen-doped graphene as the precursor.

Another interesting phenomenon with graphene is how its physical deformation affects its chemical activity. When crumpled, the carbon atoms along a fold are more reactive³. One can — at least in principle — tune the reaction barrier and reaction energetics by manipulating physical form. Very well-defined ripples or edges with curves of a few nanometres and spaced apart by, say, tens of nanometres would activate select rows of carbon atoms to achieve chemical patterns. That type of control over reactive sites could be important in nanoelectronics and nanoelectromechanical systems.

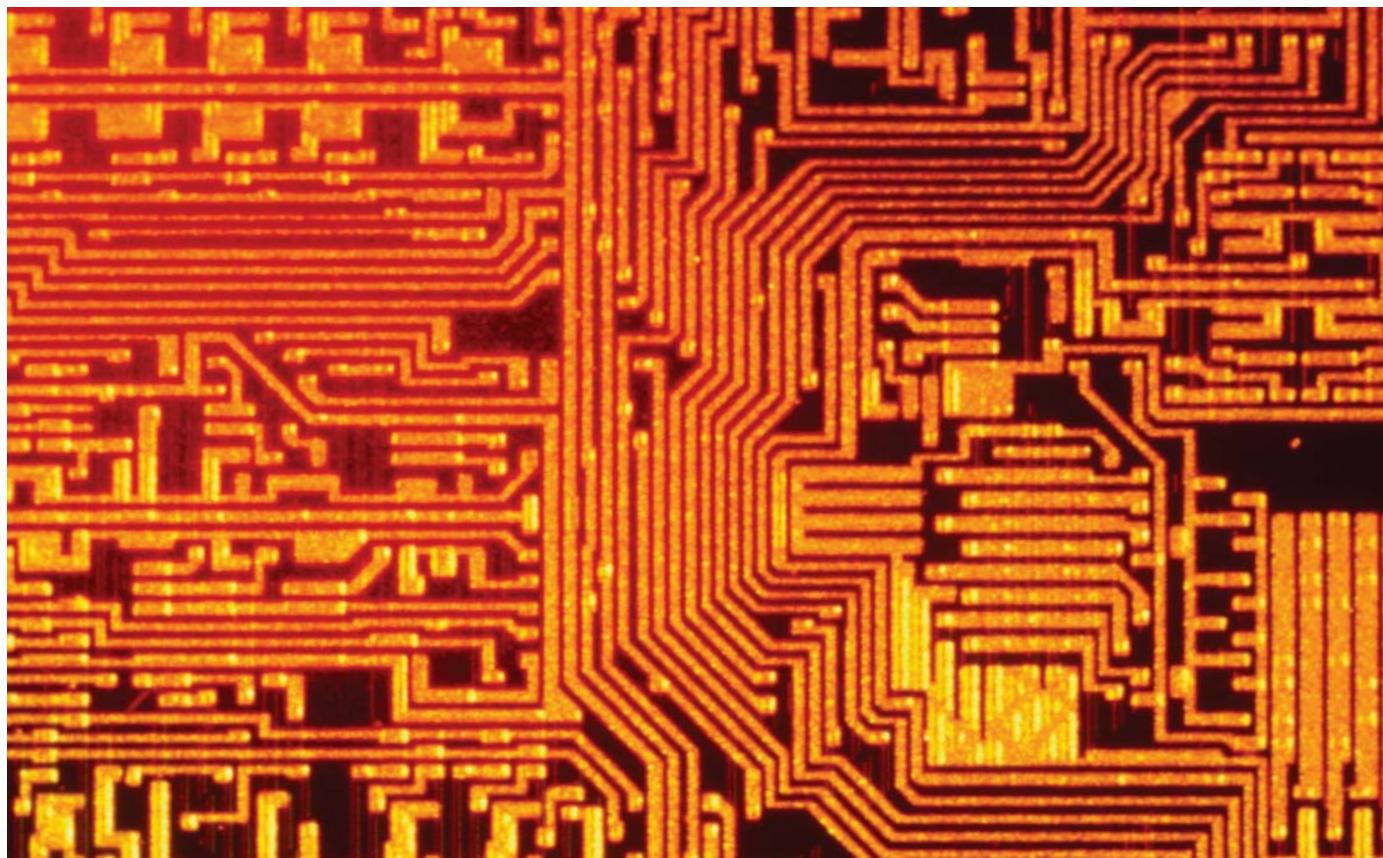
Depositing graphene on substrates patterned by nanoimprint lithography to create rows of highly curved regions would thereby allow a particular chemical function in certain regions. One could, for example, create insulating zones that border highly conductive pristine graphene zones to produce graphene nanoribbons. Alternatively, highly curved regions could be patterned by deposition onto arrays of tiny dimples, and then etched away to render an array of very small holes to make a nanomesh for filters and other applications.

Graphene platelets also deform naturally. Noteworthy forms of crumpled graphene are the reduced graphene oxide (rG-O) sheets embedded in a polymer⁴ and more recently

made⁵ by drying aerosol droplets. If this crumpling leads to localized highly reactive regions, we might find that its durability is influenced by certain chemical environments — and also that it would work well for certain applications when mangled. ■

Rodney Ruoff is professor of mechanical engineering and materials science and engineering at the University of Texas in Austin.
email: r.ruoff@mail.utexas.edu

1. Zhu, Y. *et al.* *Science* **332**, 1537 (2011).
2. Dikin, D. A. *et al.* *Nature* **448**, 457–460 (2007).
3. Srivastava, D. *et al.* *J. Phys. Chem. B* **103**, 4330–4337 (1999).
4. Stankovich, S. *et al.* *Nature* **442**, 282–286 (2006).
5. Luo, J. *et al.* *ACS Nano* **5**, 8943–8949 (2011).



MATERIAL HISTORY

Learning from silicon

Silicon is more than an incumbent technology competing with graphene — it also has a history researchers should remember.

BY MICHAEL SEGAL

Every new technology faces an incumbent, and in electronics it is usually based on silicon. Because of its intrinsic properties, its abundance and the ease with which it can be processed, silicon has dominated the semiconductor industry. It would take remarkable optimism to believe that a new material could supersede it.

Yet graphene has inspired such optimism. Its unique electrical characteristics, thinness and processability have prompted speculation that it could be the next silicon (see 'Back to analogue', page S34) — a tall order indeed. But whether or not graphene succeeds, the comparison to silicon is relevant in another sense: silicon's history gives insight into what to expect from graphene.

First, silicon teaches us that it's hard to predict the strongest or most valuable use for a novel material. Although most people associate silicon with electronic and optical devices,

that was not where the material first made its mark. Back in the 1890s, silicon was a 'poor man's alloying agent'¹ used in steel and aluminium metallurgy, competing with nickel, chromium and manganese. In the 1930s, US chemical giant DuPont intensively investigated silicon as an alternative to lead-based pigments in white paint.

These early uses were unrelated to the device in which silicon was to have its most profound impact — the transistor — but were essential for that technology's eventual development. That's because each application led to methods for producing large volumes of highly pure silicon. Research into silicon electronic devices began in the late nineteenth century with the discovery that electrical current between electrodes attached to silicon and other semiconductors flowed more easily in one direction than the other. This led to the development of semiconductor rectifiers, devices that convert alternating current into direct current, which were used for communication and radar

in both world wars. However, because only impure and granular semiconductors were available, the rectifiers were unreliable and hard to manufacture. Bell Labs and other organizations applied or adapted the production methods developed to purify silicon for metallurgy and paint to crystal rectification. This convergence was instrumental in the development of the first germanium transistor in 1947, and the first silicon transistor in 1954.

Something similar is going on now with graphene. The Nobel-prizewinning work on graphene in 2004 included an electronic application — a transistor — and electrical devices were an early focus of research². Since then, however, an increased understanding of graphene's mechanical, thermal and optical properties has broadened the scope of potential applications to include many in which graphene is not an actively controlled electronic component. These

NATURE.COM

For some of the latest graphene research: go.nature.com/no7lx5

include conductive inks and composites, transparent or high-surface-area electrodes, impermeable membranes, nonlinear optics, and nanoscale pores.

Some of these applications have spurred improvements in the way graphene is manufactured, as was the case with silicon and metallurgy. For example, the most cited graphene publication in the past two years describes not a device but rather a way to make large pieces of graphene for large-area transparent electrodes³. But many of the papers that cite this work apply the technique to devices in which graphene is an active electronic component. Similarly, one of the first mass-market applications of graphene is likely to be a conductive ink whose performance depends on a unique processing approach.

As with silicon, then, early graphene applications do not rely on its more advanced electrical characteristics, and are closely tied to steps forward in material fabrication and processing that are being adopted by the broader community. And, as with silicon, these early applications may turn out not to be graphene's greatest hit.

SWITCH HITTER

Silicon has achieved its biggest success as an electronic switch. More than a million trillion transistors are made each year⁴. But the story of the switch goes beyond silicon: the development of an electronic switch capable of amplifying a signal precedes the silicon transistor by about 50 years. The vacuum tube triode was developed around 1906 and by the 1940s was found in radios, TVs and computers. And whatever technology brings, the switch should remain a central feature: digital information is created and processed with switches, after all, and even our brains rely on a sort of switch.

Therefore, even with all of graphene's potential applications, an electronic graphene switch has special status. And, at least on first blush, it appeared that the next great switch might run on graphene. Much of the initial optimism was based on graphene's remarkable charge mobility, which is about a thousand times higher than that of silicon. Although this mobility has led to high-performance radio-frequency devices⁵, it may not be particularly useful for digital logic⁶. Moreover, graphene is missing a key characteristic required for building digital logic using traditional transistor designs: a bandgap (a range of energies in which electrons cannot exist). Already, other 2D crystals such as MoS₂, which does have a bandgap, are nipping at graphene's heels⁷. The situation recalls the competition between silicon and materials such as germanium, which has higher mobilities, and gallium arsenide, which is better suited to optical applications such as lasers. Despite these disadvantages, silicon dominated because it is easier to process, has a better oxide, and is more abundant. Graphene cannot depend on a single characteristic — electron mobility — for it to succeed as a material.

Like silicon, however, graphene brings with it a host of new physics that may change our understanding of the relevance of mobility and bandgap for transistors. Scientists are exploring alternative device architectures that exploit this new physics, including tunnelling transistors⁸ and transistors that guide and focus electrons like light.

And switches may turn out not be an important part of graphene's story after all. Kroemer's Lemma of New Technology occasionally comes up in discussion of graphene. This principle, named after Nobel laureate Herbert Kroemer, states that the primary application of a new technology is one created by that technology. If it holds here, graphene's killer app may be something that nobody has yet imagined.

GRAPHENE INKS SOME PROFITS

Although the future may be murky, graphene is expected to make its consumer product debut in 2012. MeadWestvaco, a Fortune 500 packaging manufacturer based in Richmond, Virginia, is using graphene-based conductive inks to make a new kind of package with an

Just a couple of years after the Nobel Prize was awarded, graphene is already on store shelves.

integrated anti-theft security system. The technology enables 'open merchandising', in which products are readily available for consumers to hold or see up-close. Open merchandising has traditionally been

associated with heightened security risks, particularly for high-value items packaged in paperboard or plastic.

MeadWestvaco's solution is to use graphene-based inks to monitor whether the package has been tampered with. The inks form a conductive trace that winds around the inside of the package and sticks to the product itself. A plastic tab attached to the top of the package contains powered circuitry connected to the trace; if the packaging is cut, or the tab or product removed, an alarm built into the tab sounds. Carrying the package out of the store also triggers the alarm.

Previous attempts at secure packaging like this had foundered because the right conductive ink couldn't be found. Polymer and traditional carbon-based inks are not sufficiently conductive; silver inks are expensive and harder to process. The graphene inks (supplied by Vorbeck Materials of Jessup, Maryland) work well because they contain mostly single-layer graphene sheets that are deliberately wrinkled, preventing them from re-stacking into a graphite-like material. Instead, the sheets create a network of pathways for current to flow, resulting in films that are highly conductive even at thicknesses of a few hundred nanometres. Such thin, flexible layers can be printed using newspaper-style roll-to-roll presses that are capable of high-volume production. David W. Miller, global director of security packaging

systems and supply chain at MeadWestvaco, says that major retailers have expressed an interest in deploying the packaging — and that one of them has already started testing it. Just a couple of years after the Nobel Prize was awarded, graphene is already on store shelves.

As was the case with metallurgical silicon in the nineteenth century, materials processing competence is allowing graphene to compete with established materials on this first foray into the mass market. Volumes are expected to be high, stimulating further improvements in manufacture. Vorbeck has indicated that almost all of the revenue it earns from its inks will be reinvested into research and development efforts to broaden the ink's appeal by, for example, increasing its conductivity to compete with silver inks used in solar cells. And Vorbeck's graphene inks leave the material's more exotic properties (Dirac fermions, valley polarization and so on) for future products to exploit.

ARE WE THERE YET?

So graphene is in the process of leaving the laboratory. But will we have to wait as long as we did with silicon — about 130 years from first isolation to the transistor — for graphene's golden age to arrive? Even if we must, we are already decades along the graphene development timeline. Thin graphite was being studied experimentally in the 1960s, and calculations on graphene were being performed earlier still. Much of the physics, chemistry and engineering of graphene are similar to that of nanotubes and fullerenes. In fact, many of the physicists who now study graphene have moved from research programmes on nanotubes — something reflected in the drop in growth of nanotube research papers just as the volume of graphene papers shot up after 2004.

At the same time, we are in the early or middle stages of graphene synthesis and processing. Key techniques, such as large-area chemical vapour deposition, have been developed in the past four or so years and continue to evolve (see 'Beyond sticky tape', page S32). And the number of potential applications continues to increase. Silicon's history teaches us to prioritize processing, allow for surprises, be patient, and keep a guarded (and resilient) optimism. Things are, after all, off to a good start.

Michael Segal is a senior editor of Nature Nanotechnology.

1. Seitz, F. *Electronic Genie: The Tangled History of Silicon* (University of Illinois Press, 1998).
2. Novoselov, K. S. *et al.* *Science* **306**, 666–669 (2004).
3. Bae, S. *et al.* *Nature Nanotech.* **5**, 574–578 (2010).
4. Moore, G. E. *Moore's Law at 40* (ed. Brock, D. C.) (Chemical Heritage Foundation, 2006).
5. Lin, Y.-M. *et al.* *Science* **327**, 662 (2010).
6. Schwierz, F. *Nature Nanotech.* **5**, 487–496 (2010).
7. Radisavljevic, B. *et al.* *Nature Nanotech.* **6**, 147–150 (2011).
8. Britnell, L. *et al.* *Science* (published online 2 February 2012) <http://www.sciencemag.org/content/early/2012/02/01/science.1218461>.